Journal of the
OPTICAL SOCIETY
of AMERICA

VOLUME 37, NUMBER 7 JULY, 1947

Size, Shape, and Contrast in Detection of Targets by Daylight Vision. I. Data and Analytical Description[1]

EDWARD S. LAMAR,[2] SELIG HECHT, SIMON SHLAER, AND CHARLES D. HENDLEY
Laboratory of Biophysics, Columbia University, New York, New York

(Received March 10, 1947)

The influence of size and symmetry has been studied on the contrast required for the recognition of rectangular targets against background brightnesses of 2950 and 17.5 footlamberts. Targets less than 2 minutes in diameter require the addition of a constant total light flux to the background. Larger targets require less contrast but more total flux as the area increases, until beyond 200 square minutes when the required contrast becomes independent of area. For areas below 100 square minutes, square targets are most efficient for their area; the greater the ratio of length to width the greater the contrast required. All the measurements can be unified on the supposition that the visually critical region of a target is a ribbon just inside its perimeter and about 1 minute wide. Evidently, contrast is not judged over the area of a target, but across its boundary.

I. NATURE OF INVESTIGATION

OBJECTS become visible when they differ perceptibly in brightness from their background. If I is the brightness of the background, and ΔI is the brightness difference between target and background required to render a target visible, then the Weber fraction $\Delta I/I$ is a measure of the perceptible contrast between target and background. The value of $\Delta I/I$ may be positive or negative depending on whether the target is brighter or darker than its background.

The extensive researches on contrast discrimination (for reviews see Hecht, 1924 and Steinhardt, 1936) have been mainly concerned first with establishing the relation between the just perceptible contrast fraction $\Delta I/I$ and the background brightness I; and second, with understanding the nature of the phenomenon (cf. Hecht, 1935, 1937). In general these researches were restricted to a limited number of targets and shapes, largely chosen for convenience and appropriateness.

It is common knowledge that the contrast necessary for seeing an object varies with its size. The precise relation between contrast and size has, however, not been established quantitatively, though several researches have been made for this purpose (Lasareff, 1911; Heinz and Lippay, 1928; Cobb and Moss, 1928; and Steinhardt, 1936; see also the recent measurements reported by Blackwell, 1946). Moreover, it

[1] This work was done under a contract, recommended by the Committee on Medical Research, between the Office of Scientific Research and Development and Columbia University. In addition to the detailed reports to CMR (OSRD) this material has been presented to the Optical Society of America at its meetings in October, 1946 (J. Opt. Soc. Am. **36**, 714 (1946)).

[2] Member of the Operations Evaluation Group, Navy Department. The original impetus for this research came from Dr. Jacinto Steinhardt of this group, and Dr. Steinhardt assisted at the early stages of construction and calibration of the apparatus.

is not known how the shape of a target influences the contrast required for seeing it. Since such information was needed in a variety of operational problems of military interest, we undertook the present research to supply it.

Our experiments were concerned with determining the value of $\Delta I/I$ for a series of rectangles, differing in area from 0.5 to 800 square minutes of visual angle, and with finding how $\Delta I/I$ changes for a given area of rectangle when the length/width ratio of the rectangle is varied between 2 and 200.

In order to ascertain how the results might be influenced by the prevailing brightness of the background we worked first at the high brightness of 2950 footlamberts, approximately equivalent to sunlit sand, water, and sky; and then at the low brightness of 17.5 footlamberts, roughly equivalent to the ambient brightness after sunset. Since the essential relations among $\Delta I/I$, target area, and symmetry turned out to be the same at these two extremes of daylight brightness, we confined our attention exclusively to them.

In this first paper we shall record the measurements and their analytical description. It is fortunate that even on the analytical level considerable simplicity may be achieved in presenting the variety of data secured. In the subsequent second paper we shall show how these measurements may be understood quantitatively in terms of the quantum nature of light and the mosaic arrangement of receptor elements in the retina.

FIG. 1. Observer's visual field. The observer sees monocularly a 30° circular area evenly illuminated. In the middle, delimited by a dark line, is a 4° circular area of the same brightness I as the surround. In the center of this smaller field is a black orientation dot, and near its periphery are four orientation numbers. The target appears around the central dot, and is a rectangle whose brightness above the background is ΔI.

II. GENERAL DESCRIPTION OF EXPERIMENTS

The purpose of the experiments is to measure $\Delta I/I$ for a variety of sizes and shapes of the test-object. The apparatus for this purpose is designed (1) to present the observer with a circular field 30 degrees in diameter evenly illuminated at a fixed brightness to which his eye remains continuously adapted; and (2) to add to this field at regular intervals and for exposures of 3 seconds each a rectangle of light superimposed on the background field. The added rectangle can be varied as to its (a) brightness, (b) area, (c) length to width ratio, and (d) location immediately around a central reference point.

For the observer who makes the visual measurements the problem is reduced to a simple routine. Figure 1 shows what the observer sees as he looks with his right eye through the 2 mm pupil of the apparatus. He sees the large, evenly illuminated, 30 degree circular field. In the middle of it he sees a thin dark line forming a circle 4 degrees in diameter. Immediately within the circular line 4 numbers are marked to indicate the up, right, down, and left locations of the brighter target which is to appear somewhere around a black dot in the exact center of the field. At regular intervals he hears the shutter open, and he looks for the spot of light around the central dot. He announces where he sees the spot of light. The dot is for orientation, not for fixation.

The operator who controls the apparatus sets the position of the brighter rectangular target, alters the value of $\Delta I/I$ according to a prearranged schedule, and records the results. For each selected value of $\Delta I/I$ he gives the observer 8 opportunities to say where the bright spot appears in the field, and from a plot of the relation between frequency of correct answers *versus* $\log(\Delta I/I)$, he finds the threshold value $(\Delta I/I)_t$ corresponding to 5 correct answers out of 8 judgments.

The main measurements were made with direct foveal vision. They were repeated in part with

off-axis vision 1.25 degrees in the periphery and 10 degrees in the periphery.

In the following four sections (III, IV, V, and VI) the details of the apparatus, calibration, and procedure are described for the record and for the interested reader. The reader who is concerned with the results alone is advised to skip these and to go at once to section VII.

III. APPARATUS

The apparatus is shown in diagram in Fig. 2. It is modified from an assembly designed and used earlier by Shlaer (1937) for some studies of visual acuity.

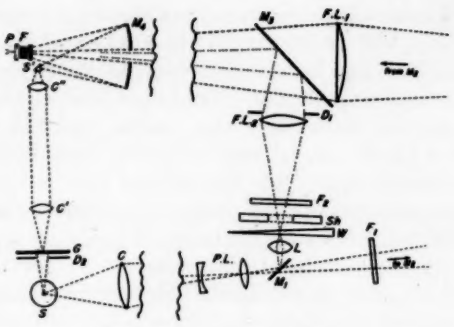


FIG. 2. Diagram of apparatus.

The source, S, is the incandescent ball of a tungsten arc. An image of the source is formed at the first lens PL by the condensing lens C. Beyond the first lens PL, the beam is divided, part being reflected by the half silvered mirror M_1 and part transmitted. The transmitted beam is reflected by the mirror M_2 (omitted in the diagram), passes through the field lens FL_1 and, after passing through the half silvered mirror M_3, is brought to a focus at the 2 mm artificial pupil P. Provision is made at F_1 for the insertion of a neutral glass filter in case it is desired to reduce the light flux in the beam just described. The eye, looking through the artificial pupil, sees this beam as a uniform visual field 4 degrees in angular diameter. This field is the circular region in the center of the 30 degree background field which the observer sees; its brightness is I, and it is referred to as the I field.

The part of the original beam which is reflected at M_1 passes through the lens L, the compensated neutral glass wedge W, the shutter Sh, the lens FL_2, and the target opening D_1 which defines the shape of the beam. At the position F_2, provision is made for inserting a neutral glass filter. Beyond the target diaphragm D_1 the beam is reflected in part at the half silvered mirror M_3 and forms an image of the source at the pupil P.

At the position D_1, provision is made for using either of two methods for producing the targets. The first arrangement is used to produce the smaller targets. It consists of two variable slits crossed at right angles so as to define a rectangular opening. The narrower of the two is a Schmidt and Haensch slit and is mounted with its fixed jaw along a radius normal to the optical axis of the system at the position D_1. The wider of the two slits was designed and constructed in our own laboratory. Its fixed jaw, the one nearest the axis, is normal to the other slit and is placed 1 mm from the optical axis. The whole slit assembly can be rotated as a unit so as to occupy any one of eight positions about the axis. When the shutter is open, the target formed by the two slits can be seen as a bright rectangle superimposed on the background field I. Since the slits do not possess sufficient range to take care of all target sizes and shapes desired, provision is made for using diaphragms with fixed openings instead of the slits. The field presented to the eye by the beam which passes through the rectangular opening forms the target in the observer's field of view. The brightness which it adds is ΔI and is controlled in steps, by the filter at F_2 and continuously by the wedge W.

An additional beam of light from the source S passes through the iris diaphragm D_2 to a ground glass G which is focused at the right angle prism S' by the condensing system C' and C''. This beam, after reflection at the prism S', fills the spherical mirror M_4 and is also brought to a focus at the pupil P. This beam appears to the observer as a uniform surround field having an angular diameter of 30 degrees. The iris diaphragm D_2 provides means of matching the surround field with the 4 degree centrally located background field I. The neutral filters F are immediately behind the pupil P and permit a brightness range, for all fields, of 8.6 logarithmic units in steps of 0.3 unit.

The optical paths from target D_1 to pupil and from lens FL_1 to pupil are the same and are equal

to 1 meter. If the observer focuses on the target, he is also in focus for objects on the surface of FL_1. This fact is made use of in providing reference marks for the observer. As a reference point to aid in determining the angular position, a small black spot is provided on the axis at the surface of the lens FL_1. Eight target positions are employed, four of the eight being marked for easy reference at the edge of the lens FL_1.

For the peripheral measurements an off-axis fixation point is provided. This spot is a small black ball situated on the end of a fine stiff wire also painted black. This is placed between the pupil and the mirror M_4 at such a position that the mirror M_4 forms an image of the ball in the plane of the lens FL_1. The ball is so oriented that its image, larger than the ball, obscures the ball from view.

IV. CALIBRATION AND ADJUSTMENT

The calibration of the filters had been done in connection with the earlier work by Shlaer (1937) and has been described by him. The wedge was calibrated with a Martens photometer.

The source was always run at 2 amperes so that its intensity remained constant. Checks of its constancy were made from time to time.

The calibration of the brightness scale was made on the 4 degree background I field. The shutter in the ΔI beam was closed and the 30-degree surround obscured by a cover over the iris diaphragm. In order to obtain a comparison field whose brightness could be measured easily for matching with the 4 degree I field, a sheet of white cardboard was erected just behind the mirror M_4, and out of the way of the opening. This cardboard was made visible to the observer by placing a mirror in front of FL_1 in such a way as to obscure half the I field. The observer was presented, therefore, with a circular field split in the middle, one-half being the I field and the other the cardboard. With a known filter at F_1, the illumination produced on the cardboard by an auxiliary light source was so adjusted as to produce a brightness match. The brightness of the cardboard was then measured by means of a Macbeth illuminometer. With the constant so determined and the known calibrations of the various filters, the brightness of the I field was then known for any combinations of filters.

The brightness of the ΔI field was obtained by comparison with the I field in the following way. The target assembly was removed so that the ΔI field had its maximum circular diameter. The appearance to the observer at this stage was of a 4-degree field of brightness I, on which was *superimposed* a central approximately 2 degree circular field from the ΔI beam. The precise part of the I field which coincided with the circular ΔI field was then blocked out by a carefully made metal disk located near the lens FL_1. The final photometric field so produced was made up of an inner circular field of the ΔI beam alone and an outer circular field of the I beam alone. The photometric field was matched, using the neutral wedge in the ΔI beam to make the adjustment. Knowing the brightness value of the I field and the match point on the wedge, the brightness of the ΔI field could be computed for any combination of filters and wedge settings.

The matching of the 30-degree surround with the 4-degree background I field was done by adjusting the iris diaphragm.

V. CORRECTION FOR DIFFRACTION AT TARGET

The purpose of the crossed slits and of the fixed diaphragms, shown at D_1 in Fig. 2 and described in Section III, is to furnish a number of rectangles of different area and symmetry in the ΔI beam so that they will appear to the eye at the pupil P as superimposed on the background field I. Because of the Maxwellian view and the mirror M_3, the eye really sees the evenly illuminated surface of lens FL_2; and the various slits and diaphragms are designed merely to control the extent and shape of this surface which is visible to the eye without altering the brightness of the surface.

Actually, however, while this is true for the larger openings, it does not hold for the very small openings because of diffraction. For small openings the converging beam which impinges on the target at D_1 changes its character on passing through the opening so that some of the light is bent out into the geometrical shadow. The resulting loss of light from the ΔI beam directed toward the pupil P causes the apparent brightness of the delimited target area on the surface of lens FL_2 to decrease. Thus a small opening not merely fixes the size of the target area visible on

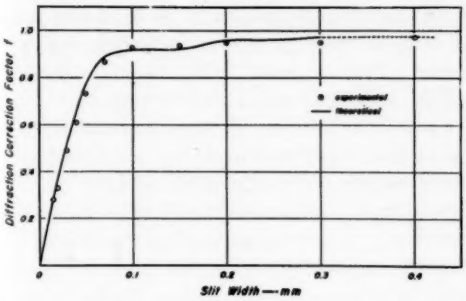


FIG. 3. Diffraction correction of target slit. For very small or narrow targets, diffraction distorts the expected brightness of the target at the observer's pupil. The results of measurements and computations for this source of error are shown.

the surface of the lens FL_2, but it also reduces the apparent brightness of the lens surface.

It is necessary to know the precise extent of the brightness reduction due to diffraction and to correct for it, else the apparent values of ΔI secured from the wedge positions and filters will not be the real ones. We determined the necessary corrections experimentally and compared them with computations made from diffraction theory.

The light flux from the ΔI beam, which passes through the pupil P, is determined by the light intensity of the beam (flux transmitted per unit area of lens FL_2 surface) and the exposed area through the target. For a fixed target area the apparent brightness of the target varies directly with the light flux through the pupil. Similarly, for a fixed intensity of the ΔI beam and for large targets—that is, those for which *both* dimensions are large and the effect of diffraction is small—the flux passing through the pupil is directly proportional to the target area. Once the correspondence between light flux through the pupil P and apparent brightness has been established for any large target, the flux to be expected in the absence of diffraction can be computed from the calibrations described in Section IV for any other target area and any brightness in the ΔI field. The ratio of the actual flux to that expected in the absence of diffraction constitutes the diffraction correction.

To limit the flux measurements to the ΔI beam alone, we covered the mirror M_2 and the iris diaphragm D_2 so as to eliminate the I field and the large surround. An RCA 929 photo-cell was mounted so as to collect the light flux passing through the artificial pupil. A d.c. balanced amplifier of the type described by DuBridge and Brown (1933) was employed as the measuring instrument.

When the experiment was tried, using the highest brightness available, we found that the light flux from the smaller targets was insufficient for accurate measurements. We decided, therefore, to measure the effect of diffraction in one dimension only, keeping the other dimension large, and to use the following approximation for the correction for any given target.

Let f_1 be the correction factor for a slit of infinite length and of width equal to the width of the target; and let f_2 be the corresponding correction for the other dimension of the target. The approximation consists in setting the total correction f equal to the product of f_1 and f_2. This approximation, although not exact, provides the first order correction whose precision is quite adequate for the purpose of this paper.

The results of the measurements secured in this way are presented as circles in Fig. 3. It is to be noted that the precision of the measurements is adequate to show not only the general trend of the curve, but also the periodicity usually exhibited by diffraction phenomena.

Adequate as these measurements appeared to be, we nevertheless decided to compare them with computations from diffraction theory. The main reason for this was the difference in spectral sensitivity between the photo-cell and the human eye. As given by the manufacturers (RCA) the spectral sensitivity of the photo-cell has its maximum farther toward the blue than that of the human eye. It is, therefore, possible that the diffraction correction appropriate for the photo-cell might be less than that for the human eye, even though the spectral energy distribution of the source is such as to offset, in part, the difference between the two detecting instruments.

The calculation of the diffraction correction for a slit of infinite length, although laborious in execution, is relatively simple in principle. An infinitely long slit is, of course, a mathematical fiction which is never realized in practice. Calculations based on this concept apply to the physical situation, provided the slit is so long that the

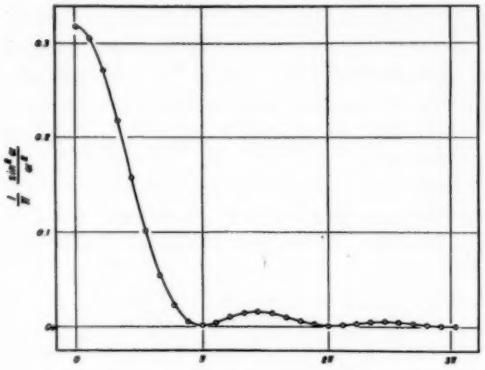


FIG. 4. Light distribution in the diffracted image of a point source.

effect of its ends can be neglected. The plane of the pupil is the image plane of the source so that the simple diffraction formula applies.

Let e be the width of the slit; λ the wave-length of the light; θ the angle in the plane normal to the slit, between the line from the source to the image plane through the slit and the line joining the slit and the point in the image plane considered; let $\alpha = (\pi e/\lambda) \sin\theta$ and f the fraction of the light flux which would be obtained in the absence of diffraction. Then for a point source of light

$$f = (1/\pi) \sin^2\alpha/\alpha^2. \qquad (1)$$

A plot of this equation is presented in Fig. 4. If in the image plane a small area is selected at an angular distance θ from the optical axis, the fraction of the undiffracted light flux passing through this area is proportional to the ordinate of Fig. 4 at the value of α corresponding to the given value of θ.

The actual source in the apparatus, instead of being a point is finite in extent, the diameter of its image at the artificial pupil P being 11 mm. To see the effect of such a finite source, let us suppose that the theoretical point source is spread out into a line parallel to the slit. Since we are considering a slit infinite in extent in this dimension, f does not change. Now suppose this line source is spread out normal to the slit. Under these conditions each strip of the source contributes to the pupil a quantity of flux proportional to the ordinate of Fig. 4 at the value of α which corresponds to its value of θ. Hence the total flux through the pupil is proportional to twice the integral under the curve of Fig. 4 over the half-width of the source as measured in units of α.

The actual source is, of course, not monochromatic so that the calculation described in the preceeding paragraph must be done for various wave-lengths. The value of f obtained at each wave-length is then weighted by the intensity of the source at that wave-length and by the photopic spectral sensitivity of the eye. The final value of f, for each slit width, is the weighted average over all wave-lengths. These theoretical values of f are presented as the solid line in Fig. 3. The good agreement between theory and experiment is apparent, and one may with confidence incorporate the diffraction corrections into the calibrations of the apparatus.

VI. PROCEDURE

The measurements are made with the fovea and with two parafoveal regions situated 1.25 and 10 degrees temporally on the retina. The procedures for foveal measurements are somewhat different from the parafoveal ones.

For the foveal measurements the field appears to the observer as shown in Fig. 1 and described in Section II. The central black dot is for orientation and for focusing. The subject knows that when a target is exposed it will appear in the immediate region of this dot in one of eight positions. He also knows when the exposure begins by the sound which the shutter makes. He therefore examines the region of the central dot and reports the position of the target. If he is not sure of the location, he makes a guess.

Each target is exposed eight times for a given brightness of the ΔI field, and the number of correct answers is determined. Since the background brightness is held constant, this measures the frequency of correct answers for a given target for a specific brightness contrast $\Delta I/I$. For each target area and shape the number of correct answers out of eight exposures is studied as a function of $\Delta I/I$, using any number of fixed values of $\Delta I/I$ that seems necessary. This may be four, or five, or even six. That contrast for which five correct answers are obtained out of eight exposures is taken as the just perceptible or threshold contrast and is called $(\Delta I/I)_t$.

The raw data from which the threshold con-

trast is secured are in some cases consistent, in others erratic as is illustrated in Fig. 5(a) and Fig. 5(b), respectively. There the number of correct answers out of eight exposures is presented as a function of the setting of the optical wedge in the ΔI beam which determines the contrast. A line is drawn through each set of data, and the wedge setting for five correct answers is selected from the line. From the wedge setting and the known calibration of the apparatus, the threshold contrast $(\Delta I/I)_t$ is determined.

This whole procedure is gone through for two background intensities of 2950 and 17.5 footlamberts; and at each intensity for ten target areas between 0.5 and 800 square minutes solid visual angle; and at each area for five length-width quotients of between 2 and 200.

Exposures are made at regular intervals by a timing device which opens and closes the shutter. Each exposure lasts 3 seconds and is followed by 7 seconds during which the subject merely keeps looking at the background field in order to keep his eye adapted to it.

The off-axis measurements are made with a different method, because the design of the apparatus does not permit orientation of the target in different large off-axis positions about a central fixation point. The method adopted is to place the fixation point horizontally off the optical axis of the apparatus and to keep the target always oriented with its long axis vertical. For the measurements made 1.25 degrees in the periphery, the lower edge of one of the orientation numbers on lens FL_1 is used as the point of fixation, while for the measurements 10 degrees in the periphery the fixation point is a special black spot already described in Section III.

TABLE I. Relation between threshold contrast $(\Delta I/I)_t$ and target area for different ratios (L/W) of length to width at a background brightness of 2950 footlamberts.

Area in square minutes	$(\Delta I/I)_t$				
	$L/W=2$	$L/W=7$	$L/W=20$	$L/W=70$	$L/W=200$
0.50	0.331	0.288	0.402	0.631	1.06
0.75	0.205	0.207	0.252	0.470	0.643
1.00	0.138	0.154	0.187	0.332	0.542
1.50	0.102	0.118	0.172	0.278	0.383
3.00	0.0582	0.0662	0.108	0.166	0.267
10.00	0.0321	0.0330	0.0445	0.0506	0.118
50.0	0.0167	0.0167	0.0182	0.0226	0.0443
100.0	0.0196	0.0132	0.0141	0.0189	
300.0	0.0166	0.0134	0.0117		
800.0	0.0200	0.0158	0.0109		

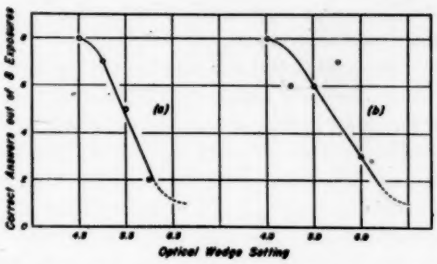


FIG. 5. Typical measurements showing the number of correct answers out of 8 exposures to a series of contrasts. The abscissas are positions of the wedge which in the apparatus controls ΔI and therefore the contrast $\Delta I/I$.

The observer is asked to look at the fixation point but to focus his attention on the target which, at the beginning, is sufficiently bright to be seen clearly. The contrast is then decreased until it is no longer visible, and the point on the wedge which corresponds to its disappearance is recorded. The average of the wedge readings for appearance and disappearance is taken as determining the threshold contrast.

After each off-axis determination of the average wedge setting, the same procedure is used to find the threshold contrast for foveal vision. This serves as a control, and the method is relied on for determining, not the absolute values of threshold contrast but the difference between the central and peripheral thresholds. Obviously the quantity which is measured using the limits method just described is not necessarily that measured using the frequency method.

In all the measurements care is taken to see that the observer gets a sufficient number of rest periods to avoid fatigue.

VII. FOVEAL MEASUREMENTS OF THRESHOLD CONTRAST

About a dozen subjects were employed in making the measurements. These were all under 30 years of age and had 20–20 vision or better. Because of circumstances beyond our control all of them did not participate in all the measurements. Four subjects completed all four groups of measurements, while five completed all but the 10-degree peripheral trials. Except for the 10-degree peripheral measurements, the results to be reported are based on the observations of those five subjects who completed the remaining sets. However, it may be said at once that no signifi-

cant differences appear if all the data on all subjects are included.

The average measurements for the fovea with the five observers are given in Table I for the background brightness of 2950 footlamberts, and in Table II for the background of 17.5 footlamberts. For easier understanding the measurements are also shown graphically in Figs. 6 and 7. The ordinate, in each case, is the threshold contrast $(\Delta I/I)_t$ while the abscissa is the angular area of the target. For both abscissa and ordinate, a logarithmic rather than a linear scale is employed to prevent crowding of the experimental points. The various curves shown in each figure are for various asymmetry quotients representing the ratio of length to width of the rectangular target.

Examination of Figs. 6 and 7 shows that there is little difference between the data for asymmetry quotients 2 and 7. Evidently for small asymmetries the ratio of length to width is unimportant; the determining variable is the area of the target, not its shape. Beyond the small asymmetries, however, target shape becomes important, and the contrast threshold rises as the target becomes narrower and longer.

This is true for the small and medium-sized targets. For the larger targets, the effect of symmetry seems to be unimportant. The lines for the different asymmetries come together and even appear to cross near 100 square minutes.

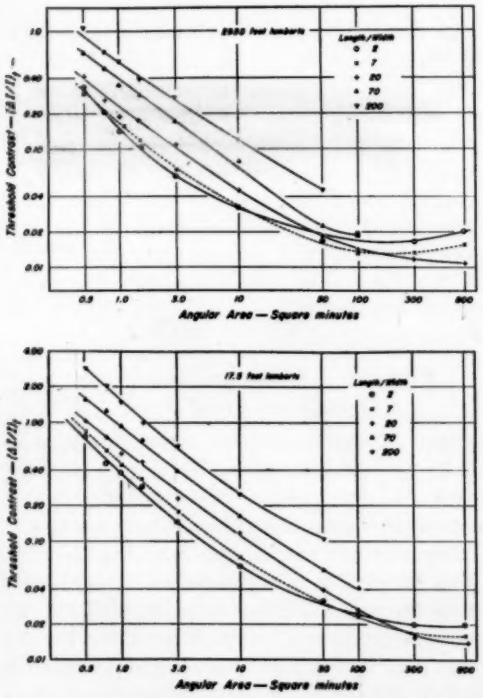
It is also to be noted in Figs. 6 and 7 that for small areas, the slope of all the lines approaches 45 degrees. This means that for areas less than about two square minutes, contrast and area are reciprocally related in such a way that one may be quantitatively substituted for the other. For a given shape of target a given total flux of light is required for the recognition of a threshold contrast, and this total flux can be made up by a change in either the brightness or the area of the target, provided the area is small.

The fact that with increasing area the slope of the lines rapidly becomes less than 45 degrees and finally reaches a value of zero at the largest areas, shows that the simple reciprocal relation between contrast and area does not hold. Beyond the smallest areas, an increase in area does not produce a correspondingly similar decrease in threshold contrast, and for the largest areas an increase in area produces no decrease in threshold contrast at all.

For two asymmetries at the high brightness in Fig. 6, the largest area actually requires an increase in contrast threshold. Whether these two points are real or the result of random errors cannot be decided without much careful experimentation; such a rise in the threshold contrast

TABLE II. Relation between threshold contrast $(\Delta I/I)_t$ and target area for different ratios (L/W) of length to width at a background brightness of 17.5 footlamberts.

Area in square minutes	$(\Delta I/I)_t$				
	$L/W=2$	$L/W=7$	$L/W=20$	$L/W=70$	$L/W=200$
0.50	0.715	0.850	1.02	1.54	2.86
0.75	0.450	0.576	0.750	1.24	1.99
1.00	0.373	0.434	0.553	0.911	1.46
1.50	0.282	0.330	0.466	0.702	0.979
3.00	0.141	0.177	0.228	0.382	0.632
10.0	0.0594	0.0716	0.115	0.161	0.246
50.0	0.0306	0.0312	0.0394	0.0572	0.103
100.0	0.0244	0.0237	0.0266	0.0408	
300.0	0.0197	0.0165	0.0154		
800.0	0.0195	0.0162	0.0141		



FIGS. 6 and 7. Relation between area and threshold contrast $(\Delta I/I)_t$ for different ratios of length to width of the rectangular target. Foveal vision.

does not appear at all in Fig. 7, and we are inclined to lay no emphasis on its appearance in the two instances in Fig. 6. In fact, the rise disappears in the quantitative transformations which the data are subjected to in the following sections.

VIII. FLUX, LENGTH, AND WIDTH

By mere inspection of the data in Figs. 6 and 7 we have been able to make some deductions as to their trends. However, the measurements in Tables I and II possess many more correlations and regularities, and these can be made apparent only by closer study. We shall make this study for the foveal measurements because of their relatively good precision and definitiveness. Nevertheless, it should be said at once that the peripheral measurements show essentially the same trends as the foveal data and are subject to the same kind of analysis.

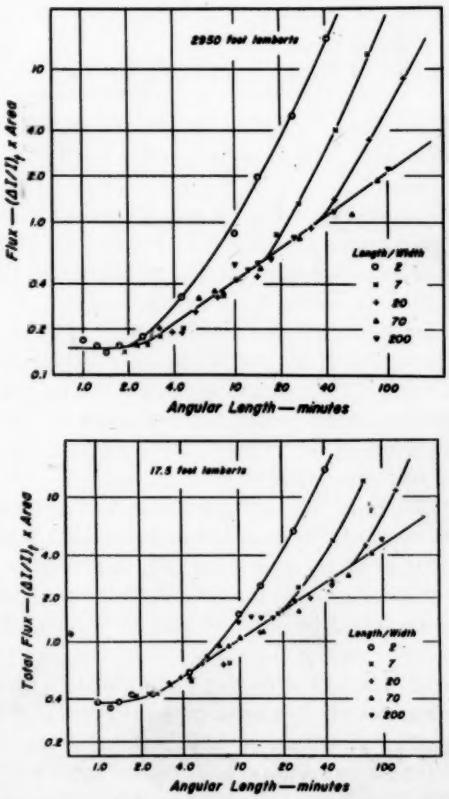
The problem of correlating all the data of Tables I and II was approached in a number of different ways. The most fruitful was to examine the relations which the total light flux needed to see a target bears to its dimensions.

The total light flux which enters the eye from any target is proportional to the product of its angular area and the threshold contrast $(\Delta I/I)_t$. We will call this product the flux F, and will relate it first to the angular length of the target, and second to its angular width. Figures 8 and 9 show the values of this flux F for the two background brightnesses of 2950 and 17.5 footlamberts, respectively. The abscissa in both cases is the angular length of the corresponding target.

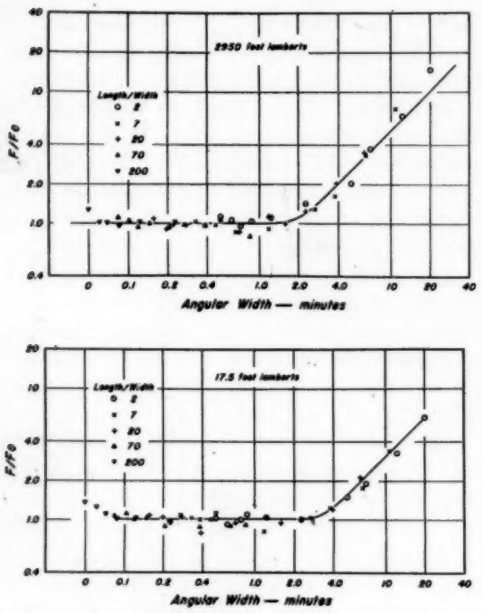
In Figs. 8 and 9 it will be noticed that the flux F required for contrast recognition is constant so long as the angular length of the target does not exceed some two or three minutes. Beyond this target length the flux required to see the target increases. At each target length there is a definite lower limit to the flux required. The line, drawn through the lower limit for each target length, describes a flux which is here designated as F_0. It will be noticed that the data for each of the various asymmetries depart, one after the other, from this F_0 line. The target width at which the departure begins is approximately the same for all asymmetries, namely, about two minutes.

Let us assume as a first approximation that the flux required for seeing a target can be represented as the product of a function of length and a function of width. Then F_0 is the desired function of length and F/F_0 is the desired function of width. In Figs. 10 and 11, F/F_0 is presented as a function of target width for background brightnesses of 2950 and 17.5 footlamberts respectively. The fact that all the points, regardless of target size or shape, lie on the same curve for each brightness indicates that the flux F required for seeing the target can be represented, for engineering purposes, as the product of a function of target length and a function of target width.

This result, though useful from an engineering point of view, is most likely fortuitous, and, in fact, it is probably misleading from the point of view of understanding the performance of the



FIGS. 8 and 9. Relation between contrast flux and angular length of target. Foveal vision.



FIGS. 10 and 11. Relation between flux F/F_0 and the width of a rectangular target. Foveal vision.

human eye. The reasons for this deserve some discussion.

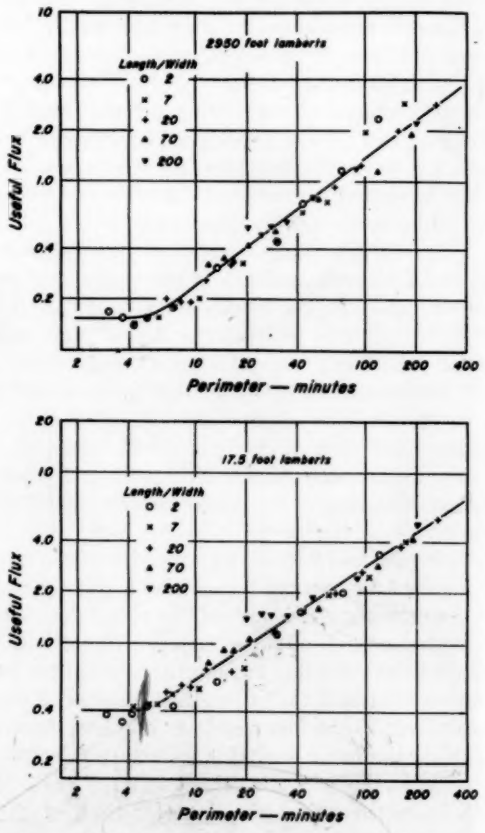
Comparison of Figs. 8 and 10 and the corresponding ones, Figs. 9 and 11, shows that beyond the two or three minutes over which F_0 and F/F_0 are both constant, the two functions are not of the same form. Hence the form of the functional dependence on the second variable presupposes the existence of a first dimension which is greater than the second. In this sense, therefore, the two dimensions of the target cannot be considered as independent variables.

While this is not an impossible situation physiologically, it is a complex one which should be explored only if no other avenue is open. It is far simpler to infer that the mathematical relation described by Figs. 10 and 11 is fortuitous, especially since a slightly different treatment of the data themselves produces a formulation which makes more obvious sense physically and physiologically. However, it has been necessary to present the two pairs of relations in Figs. 8 and 10 and in Figs. 9 and 11 so as to be able to formulate a different treatment from them.

IX. USEFUL FLUX

Going back to Figs. 8 and 9, one can see that beyond an angular length of 2 or 3 minutes the F_0 line goes up at less than 45 degrees. This means that an increase in length results in a decrease in $(\Delta I/I)_t$. In other words, the greater the length of a narrow target, the smaller is the threshold contrast required to see it.

On the other hand, one can see in Figs. 10 and 11 that beyond an angular width of 2 or 3 minutes, the F/F_0 line does go up at 45 degrees. This means that an increase in target width, beyond 2 or 3 minutes, does not result in a decrease in the threshold contrast required to see the target. In other words, when the width of a target is enlarged beyond 2 or 3 minutes, the in-



FIGS. 12 and 13. Relation between useful flux and perimeter of target. The critically useful flux is the light from an area 1 to 1.5 minutes just inside the perimeter of the rectangular target. Note how this concept unifies all the measurements.

crease in area has to be made at the same brightness as was necessary for the 2 or 3 minute width. It is as if for a width larger than 2 or 3 minutes each added area renders useless for sensory purposes a similar area which is already there. This, essentially, is saying that flux contributed by points more than 1 or 1.5 minutes in from the edge of the target does not make the target more visible to the observer. In the contrast recognition of a target, the physiologically useful flux arises from a strip inside the target which contains all the points 1 or 1.5 minutes or less from the edge.

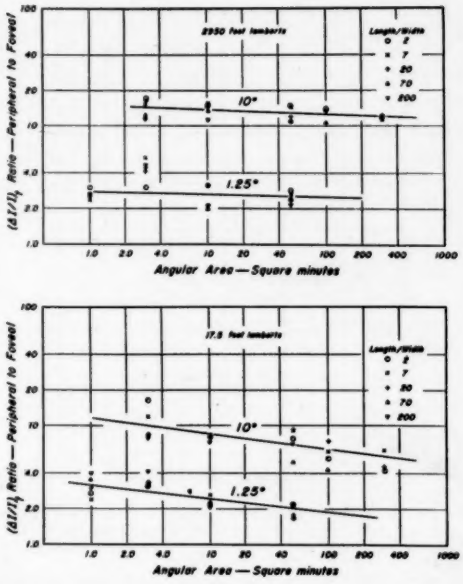
If this concept of physiologically useful flux has meaning, the magnitude of useful flux should bear a simple relation neither to the width nor to the length of a target, but to its perimeter. This is indeed true. We have computed the useful flux in all the targets and present them in Figs. 12 and 13 in relation to the perimeter of the targets. The limiting strip width employed in computing this useful flux was 1 minute for a brightness of 2950 footlamberts (Fig. 12) and 1.5 minutes for 17.5 footlamberts (Fig. 13). The difference in strip width in the two cases is entirely empirical: the data fit better. The possible reasons will be discussed in the next section.

It is clear that the concept of useful flux unifies all the data at each background brightness. With the exception of two points representing the two smallest areas of the target with the smallest ratio of length to width, all the measurements lie on a straight line. The slope of the line for the high brightness is $\frac{3}{4}$ and for the low brightness $\frac{2}{3}$. Since Figs. 12 and 13 are drawn on a double logarithmic grid, it follows that the useful flux varies as the $\frac{3}{4}$ or $\frac{2}{3}$ power of the perimeter depending on the background brightness.

Let L and W be the length and width of the target; let w be the width of the useful ribbon just inside the perimeter R. Then the useful area A_1 is equal to $2w(L+W-2w)$, and the perimeter R is equal to $2(L+W)$. The useful flux then becomes $(\Delta I/I)_t A_1$, and it varies as a fractional power of the perimeter. From the plots in Figs. 12 and 13 it follows that:

$$(\Delta I/I)_t = cR^k/A_1 \qquad (2)$$

where c is a constant corresponding to the antilog of the intercept of the line, and k is an exponent whose value is either $\frac{3}{4}$ or $\frac{2}{3}$ depending on the background brightness.



FIGS. 14 and 15. Comparison of parafoveal (1.25°) and peripheral (10°) measurements with foveal measurements of threshold contrast for different areas and symmetries of targets.

It can be seen from Figs. 12 and 13 that the concept of useful flux gives at least as satisfactory a representation of the data as does the engineering representation described earlier. It possesses the advantage of suggesting the mechanism employed by the eye in making a visual judgment. The concept of physiologically useful flux indicates that the two quantities which determine whether or not a given target can be seen are the light flux received from points within a small angular distance from the target edge, and the length of the edge. In other words, the suggestion is strong that the eye makes its judgment of contrast across a boundary rather than over an area.

X. THRESHOLD CONTRAST IN THE PERIPHERY

The method employed for the peripheral measurements and described in Section VI does not have the potentialities for accuracy of the method used for the fovea. Therefore, we relied on the measurements only to provide a comparison of

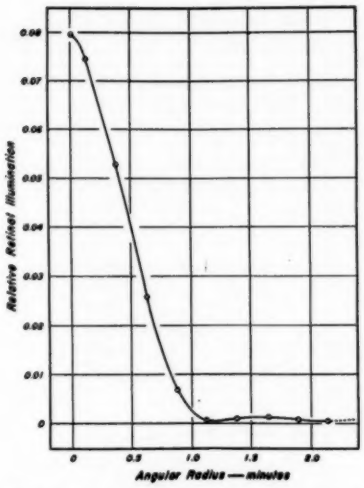


FIG. 16. Retinal illumination produced due to diffraction by a 2 mm pupil, as computed from equation (3).

the periphery with the fovea and not to yield absolute values.

Even in making the comparisons, one may well ask whether or not the same quantity is measured by the two methods. To avoid misgivings, we made comparisons between the absolute values of threshold contrast for the fovea, as determined by the two methods. Within the experimental error there was no difference, although within this random error there was a slight trend. This trend indicates greater dependence of useful flux on perimeter for the less accurate method.

The ratios of foveal and peripheral values of threshold contrast are presented in Figs. 14 and 15 for 2950 and 17.5 footlamberts background brightness, respectively. The abscissa in each case is the area of the target in square minutes. These data indicate a slight decrease in this ratio as either area or asymmetry quotient is increased. Furthermore this ratio seems to be slightly smaller at the lower brightness.

The great similarity between the foveal and peripheral data indicates that the treatment of the peripheral data in terms of the useful flux concept should result in the same values of strip widths. We have actually plotted these data in terms of useful flux as a function of target perimeter, and the same values of strip width do, in fact, provide the best representations of the data.

XI. PHYSICAL BASIS OF USEFUL FLUX

The notion that the significant sensory events in contrast discrimination take place across a boundary rather than over an area conforms to general photometric experience. In this connection we derived in Section IX the concept of physiologically useful flux and showed how the measurements all fall together in terms of it. It would now seem desirable to consider the meaning of this concept more carefully, particularly in relation to the actual physical conditions at the boundary. In order to do this, it is necessary to review in some detail the passage of light from a target through the optical system of the eye and the final formation of an image on the retina.

Consider first a target so small that it is essentially a point. Because of the phenomena of diffraction and of chromatic, spherical, and other aberrations in the optical system of the eye, this point target produces on the retina not a point image but a rather fuzzy disk several minutes in angular diameter. With the 2 mm artificial pupil employed in these experiments this fuzzing of the image over a finite retinal area is almost entirely the result of diffraction.

The distribution of retinal illumination which results from diffraction at a circular pupil has been worked out for a point target and can be expressed, for a monochromatic point source of unit intensity, by:

$$E = (1/\pi)[J_1(\alpha)/\alpha]^2, \qquad (3)$$

where J_1 is a Bessel function of the first order, α the angular variable already defined in Section V,

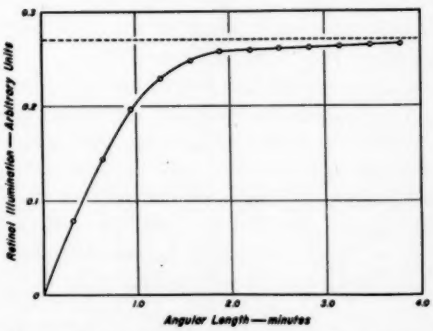


FIG. 17. Illumination at the central point of a retinal image as the size of the target increases. Only a slight increase occurs after the target dimension is greater than 2 minutes.

and E is the retinal illumination. If all wave-lengths are present, the distribution is obtained by weighting that for each wave-length by the spectral energy distribution of the source and the spectral sensitivity of the cones.

We have made this computation of Eq. (3), and the results are presented in Fig. 16. Here the retinal illumination is treated as ordinate, while the abscissa is the angular distance from the center of the image, expressed in minutes.

Starting with the image of a point target produced on the retina, let us consider what happens to the added illumination at some spot in the image as the dimensions of the target are altered while keeping its brightness the same. The most convenient spot to take for illustration is the one in the center of the image. Suppose that the point target is now drawn out into a line keeping the brightness of the target the same. As the length increases, the illumination at the center of the image increases, at first rapidly and then more slowly, approaching a constant maximum value. This is because as the length increases, each strip of the target contributes to the illumination of the central spot a quantity of light proportional to the ordinate in Fig. 16 at the abscissa value of α corresponding to the angular position of the strip. The total illumination at the center of the image is therefore proportional to twice the integral or area under the curve of Fig. 16.

This integral is shown in Fig. 17 as a function of the angular length of the target. It can be seen from Fig. 17 that the rapid change in retinal illumination is pretty well over when the target length reaches 2 minutes and that beyond 2 minutes the change in retinal illumination with increasing target length is small.

If this line target is now spread out in the other dimension, again keeping the target brightness the same, the same general effect is repeated. As the width is increased, the illumination of the center of the image increases rapidly until the width is about 2 minutes, and then more slowly, approaching a constant maximum value.

The computation of the retinal illumination can be carried out by graphical integration methods for any given point in the image of any given target. However, it is unnecessary to do this, because the discussion in connection with Figs. 16 and 17 is adequate to clarify the behavior of the measurements.

We have found that for targets of small area, the threshold contrast is inversely proportional to the area. This is understandable in terms of Figs. 16 and 17. The blurred, diffracted image of all these areas is practically constant in size, but varies in brightness as the target area is altered. In other words, a change in the area of the target can be compensated for by a reciprocal change in the added brightness. The physical result is the same; it is therefore not surprising that the physiological result is the same.

Notice in Fig. 17 that for any symmetry of target the retinal brightness at the center of the image steadily increases as either dimension of the target increases up to 2 minutes, after which it remains constant. Below 2 minutes therefore, a change in this dimension of the target produces the same effect as a change in brightness in the diffraction pattern of that dimension. This is precisely what the measurements in Figs. 8–11 state. When either the width or the length of the target is less than 2 minutes, a change in that dimension makes no change in the total flux required for seeing the target.

Finally, the concept of useful flux indicates that the critical sensory events in brightness discrimination take place in the narrow ribbon around the retinal image of the target which is comprised by the retinal brightness gradient produced by diffraction. As Figs. 16 and 17 show, this ribbon of gradient is between 1 and 2 minutes wide. Since foveal cones are about 0.5 minute in diameter, the gradient width covers about 3 or 4 cones. It is to the behavior of those cones that one must look for an understanding of brightness discrimination.

It is apparent that, even without a complete computation of the brightness distribution in the retinal image of the various targets, one can deduce a good deal about the behavior of brightness discrimination. It might be worth while to make such a complete computation. But this is a tedious procedure, and we therefore suggest a relatively simple approximation which, though not complete, is adequate to describe the present measurements.

Let us designate by E the background illumination of the retina and by ΔE the difference

between the illumination of the background and that inside the image. We are interested in expressing $\Delta E/E$ in terms of $\Delta I/I$ and the dimensions of the target. The approximation required to do this must have the following properties. First, for any given target, $\Delta E/E$ must be proportional to $\Delta I/I$. Second, keeping $\Delta I/I$ constant, $\Delta E/E$ must increase rapidly with increasing target size, provided the minimum angular dimension of the target is less than $2w$ where w is a critical angular distance to be discussed. Third, keeping $\Delta I/I$ constant, $\Delta E/E$ must increase slowly with increasing target size if the minimum angular dimension exceeds $2w$, approaching a maximum value independent of target size.

An approximation which has these properties is

$$\Delta E/E = C(\Delta I/I)(A_1/A_2) \qquad (4)$$

where C is a constant, A_1 is the angular area of that part of the target which lies within a distance w of the target edge, and A_2 is the product of w and the angular perimeter of the image.

For small targets A_1 is the whole target area, and the image perimeter is constant so that $\Delta E/E$ is proportional to target area, an excellent approximation to the result desired.

For large targets, those more than $2w$ wide, the image perimeter and the target perimeter are about the same. If L and W are the angular length and width of the target, respectively, then for large targets,

$$A_1/A_2 = [LW-(L-2w)(W-2w)]/2(L+W)w$$
$$= 1-2w/(L+W). \qquad (5)$$

This function changes slowly with L and W and approaches unity for large values of these two variables. Hence $\Delta E/E$, which is proportional to A_1/A_2, changes slowly with L and W and approaches a limiting value independent of target size. This is the behavior required for large target sizes.

An examination of Eq. 4 shows that part of it, that is $(\Delta I/I)A_1$, is simply the useful flux described in Section IX. Hence if the useful flux is divided by A_2, the resulting quantity is proportional to $\Delta E/E$, the Weber fraction expressed in terms of retinal illumination. From the computed data presented in Fig. 12 and Fig. 13, therefore, one can study the behavior of this retinal Weber fraction as a function of the perimeter of the image. The results obtained for the two brightnesses are different. The high brightness data indicate that the product of the threshold value of $\Delta E/E$ and the fourth root of the image perimeter is constant. The low brightness data indicate that the product of $(\Delta E/E)_t$ and the cube root of the image perimeter is constant.

The critical strip width w obtained from the high brightness data is 1 minute. This is in good agreement with expectation based on the assumption that diffraction is responsible almost entirely for the spreading of the light flux by the optical system of the eye. The critical strip width from the low brightness data, on the other hand, is 1.5 minute.

There may be a number of possible explanations of this difference. One of these is presented here. At high brightness, the actual pupil and the artificial one are about the same size. If the eye of the observer is not centered on the artificial pupil, he notices a marked decrease in the apparent brightness of the field and quickly recenters his eye so that the high brightness data are taken through the central part of the optical system. Here the aberrations are reduced in effect so that diffraction is the chief cause of flux spreading. At the lower brightness the actual pupil is much larger than the artificial one so that the observer may have his eye way off center without being aware of any marked decrease in apparent field brightness. Hence the low brightness data may have been taken through a part of the optical system which suffers from the various lens aberrations. The causes of flux spreading under these conditions are not only diffraction but lens aberrations as well so that the limiting strip width w can be expected to exceed 1 minute.

SUMMARY

1. Measurements have been made of the brightness contrast required for the detection of a series of rectangular targets brighter than their background at a brightness of 2950 footlamberts, corresponding to sunlit sand, water, and sky, and at 17.5 footlamberts corresponding to the ambient brightness shortly after sunset. The targets varied in area between 0.5 and 800 square

minutes and differed in the ratio of length to width between 2 and 200.

2. For the detection of targets both of whose dimensions are less than 2 minutes, the total flux added to the background by the target is constant. The threshold contrast $(\Delta I/I)_t$ varies inversely with area, and the product of the two is constant. The product of $(\Delta I/I)_t$ and the area in square minutes is 0.14 and 0.36 for the high and low brightnesses respectively.

3. The simple reciprocal relation between area and contrast breaks down when one dimension of the target is larger than 2 minutes. More total flux is needed for its detection than would be expected from the reciprocal relation.

4. For long targets whose width is less than 2 minutes, variations in width can be compensated by inverse variations in contrast.

5. Targets whose both dimensions are greater than 2 minutes become increasingly less efficient visually as they increase in size. They require more flux as the area increases until the required contrast becomes independent of area.

6. For areas below 100 square minutes, square targets (and undoubtedly circular ones) are most efficient for their area. The greater the ratio of length to width in a target the more flux does it require for detection.

7. For these large areas it appears that the visually critical region of the target is the ribbon just inside its perimeter and 1 or 1.5 minutes wide. This visually critical area multiplied by the threshold contrast is directly proportional to the $\frac{3}{4}$ or $\frac{2}{3}$ power of the perimeter of the target for the high and low brightnesses respectively. The light from this critical area has been called the physiologically useful flux. Evidently the eye makes its judgment of contrast not over the area of a target, but across its boundary.

8. Similar regularities hold for parafoveal (1.25 degrees out) and peripheral (10 degrees out) vision at the two brightnesses, except that the values for the threshold contrast are different for extra-foveal regions. The threshold contrast for parafoveal vision is about three times that of foveal vision at both brightnesses. The threshold contrast for peripheral vision is about 15 times that of the fovea at the high brightness, and about eight times that of the fovea at the low brightness.

BIBLIOGRAPHY

H. R. Blackwell, "Contrast thresholds of the human eye," J. Opt. Soc. Am. **36**, 624 (1946).

P. W. Cobb and F. K. Moss, "The four variables of the visual threshold," J. Frank. Inst. **205**, 831 (1928).

L. A. DuBridge and H. Brown, Rev. Sci. Inst. **4**, 532 (1933).

S. Hecht, "The visual discrimination of intensity and the Weber-Fechner law," J. Gen. Physiol. **7**, 235 (1924).

S. Hecht, "A theory of visual intensity discrimination," J. Gen. Physiol. **18**, 767 (1935).

S. Hecht, "Rods, cones, and the chemical basis of vision," Physiol. Rev. **17**, 239 (1937).

M. Heinz and F. Lippay, "Über die Beziehungen zwischen der Unterschiedsempfindlichkeit und der Zahl der erregten Sinneselemente. I. Mitteilung. Abhängigkeit der Unterschiedsempfindlichkeit für Helligkeiten von der Grösse des Netzhautbildes. Versuche bei fovealer Beobachtung," Arch. ges. Physiol. **218**, 437 (1928).

P. Lasareff, "Studien über das Weber-Fechner'sche Gesetz. Einfluss der Grösse des Gesichtsfeldes auf den Schwellenwert der Gesichtsempfindung," Arch. ges. Physiol. **142**, 235 (1911).

S. Shlaer, "The relation between visual acuity and illumination," J. Gen. Physiol. **21**, 165 (1937).

J. Steinhardt, "Intensity discrimination in the human eye. I. The relation of $\Delta I/I$ to intensity," J. Gen. Physiol. **20**, 185 (1936).

The Sensitivity of the Human Eye to Infra-Red Radiation

DONALD R. GRIFFIN, RUTH HUBBARD, AND GEORGE WALD*
The Biological Laboratories of Harvard University, Cambridge, Massachusetts

(Received March 12, 1947)

The spectral sensitivity of human vision has been measured in the near infra-red, in two areas of the dark adapted eye: the central fovea (cones) to 1000 mμ, and a peripheral area, in which the responses are primarily caused by rods, to 1050 mμ. In both cases the estimates of spectral sensitivity are based upon determinations of the visual thresholds for radiation passing through a series of infra-red filters. By successive approximation, sensitivity functions were chosen which were consistent with the observed thresholds.

The spectral sensitivity of the fovea determined in this way is consistent with previous measurements of Goodeve on the unfixated eye. At wave-lengths beyond 800 mμ the periphery becomes appreciably more sensitive than the fovea. This tendency increases at longer wave-lengths, so that at the longest wave-lengths studied, the radiation appeared colorless at the threshold and stimulated only rods.

Lengthening the exposure time increases the sensitivity of the peripheral retina relative to the fovea. Our measurements involved exposures of 1 second and fields subtending a visual angle of 1 degree. With shorter exposures or smaller fields the fovea is favored, so that under such circumstances the fovea may become more sensitive than the periphery well into the infra-red.

At 1050 mμ the sensitivity of the peripheral retina is only 3×10^{-13} times its maximum value at 505 mμ. A computation shows that by 1150 or 1200 mμ radiation should be more readily felt as heat by the skin than seen as light by the eye.

IT is customary to set the limits of the visible spectrum at about 400 and 750 mμ, and to consider that shorter or longer wave-lengths are invisible ultraviolet or infra-red. However, recent measurements have shown that with sufficient energy the eye can be stimulated by wave-lengths well outside of these limits. Goodeve (1936) has measured the visual threshold at wave-lengths as long as 900 mμ,[1] and has discussed the visibility of ultraviolet radiations down as low as 312.5 mμ (1934).

In this paper we shall describe new measurements in the near infra-red which extend to 1000–1050 mμ, and in which fixation of the eye was adequately controlled so that data could be obtained on the sensitivity of both foveal cones and peripheral rods. Previous work had shown that from 550 to 700 mμ the rods decline in sensitivity more rapidly than the cones and that in the far red the fovea is the more sensitive portion of the retina (Abney and Watson, 1916; Wald, 1945b). Our measurements show that at longer wave-lengths this tendency is reversed and that beyond 800 mμ, under our conditions of observation, the rods are appreciably more sensitive than the cones. At the threshold of the whole dark-adapted eye, therefore, infra-red lights are seen only with the peripheral retina, and they appear colorless.

APPARATUS AND METHODS

Our method of obtaining sensitivity functions for the fovea and periphery in the infra-red was similar to that used by Dresler (1937) in the ordinary visible spectrum. It was based upon determinations of the visual transmissions of several filters—the ratios of the transmitted to the incident luminous flux. For this purpose we measured the visual threshold of the dark-adapted eye, using each filter in turn, and varying the intensity of the stimulus by means of a neutral optical wedge.

Such thresholds are inversely proportional to the visual transmissions of the filters. The visual transmission of each filter could also be calculated from its physical properties and an assumed sensitivity function for the eye. By successive approximation the assumed sensitivity function was adjusted until the calculated visual transmissions were brought into agreement with the

* This research was performed under contract with the Army Engineer Board. An account of it was presented at the Cleveland meeting of the Optical Society in March, 1946 (Griffin, Hubbard, and Wald, 1946).

[1] A recent paper by N. I. Pinegin (1945) has been received since our work was completed. Visual thresholds were measured at 655, 700, 800, 900, and 950 mμ, using as stimulus the exit slit of a monochromator viewed through a 9× eye-piece. The angular subtense of the test field is not stated. The results agree closely with Goodeve's foveal sensitivity function, and the color of the 950 mμ radiation was reported as "an unsaturated red" at threshold, so that it seems clear that foveal vision was involved.

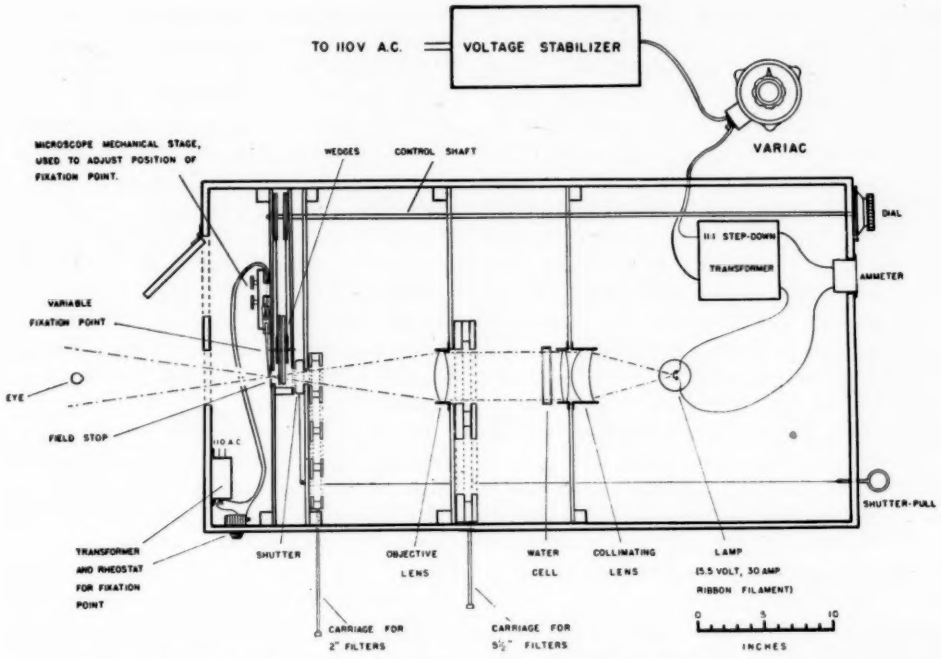


FIG. 1. Top view of adaptometer used to measure visual thresholds in the red and infra-red.

measurements. The final corrected function represents the sensitivity of foveal cones and peripheral rods from 700 mμ to 1000 and 1050 mμ, respectively (see Fig. 4 and Table V).

Our threshold measurements were made with a specially constructed adaptometer shown in schematic top-view in Fig. 1. A ribbon-filament tungsten lamp was used as source, and an image of the filament was focused on the retina. An f:1.6 lens was used to concentrate the necessary energy to enable the observer to see the filament through the very dense infra-red filters which we employed. A water cell 2 centimeters thick excluded long wave-length infra-red radiation, and provision was made for inserting the infra-red filters in a collimated portion of the light path. The light was focused after passing through water cell and infra-red filter to form an image of the filament at a field stop which could be varied in size to control the area of retina stimulated. Just before reaching this focus the light passed successively through 2-inch filters, if these were wanted, a photographic shutter to control exposure time, and a calibrated neutral optical wedge to control the intensity.

A movable fixation point similar to that described by Wald (1945a) was located very close to the plane of the field stop. For most of our measurements the image of the filament was viewed directly from a distance of 14 inches; but in some cases a magnifying ocular was used to increase the area of retina stimulated and incidentally to concentrate more radiation into the eye. A chin-and-forehead rest fixed the observer's head.

The optical wedges were made from photographic plate by graded exposure, and they permitted continuous variation of intensity over a range of about 1,000,000 times. Their construction was identical with those described by Wald (1945a) and they were calibrated similarly, by means of a Beckman spectrophotometer. The calibration was carried far enough into the infra-red to make certain that the calibration curve was accurate for all wave-lengths used in our threshold determinations.

The greatest care had to be exercised to avoid any leakage of light, either between the partitions or to the outside. The dark-adapted eye is about 10^{11} times more sensitive at 500 mμ than

at 944 mμ, the wave-length at which some of our filters transmitted their greatest luminous flux. Leaks of white light which ordinarily would be negligible may in this case be many times more effective in stimulating the eye than the infra-red radiation which the apparatus is intended to deliver.

Before making threshold measurements the observer's eye was aligned with the optic axis of the instrument by means of the adjustable chin-and-forehead rest. He was allowed to become adequately dark-adapted and was then asked to fixate upon the dim fixation point while flashes of light were presented by means of the shutter. We always approached the threshold from below, i.e., by setting the intensity initially at a sub-threshold level and raising it until the observer reported seeing a flash. Final determinations were ordinarily made with intensity intervals of approximately 0.07 log unit. Several such readings were made, and a series was considered to be satisfactory when the individual measurements showed no systematic drift and varied within a range of not more than 0.2–0.3 log unit. The value assigned to the threshold was the average of such a series of readings. Our observers ranged in age from 17 to 40 and most of them were men. All had 20/20 vision, either with their unaided eyes or with correcting lenses.

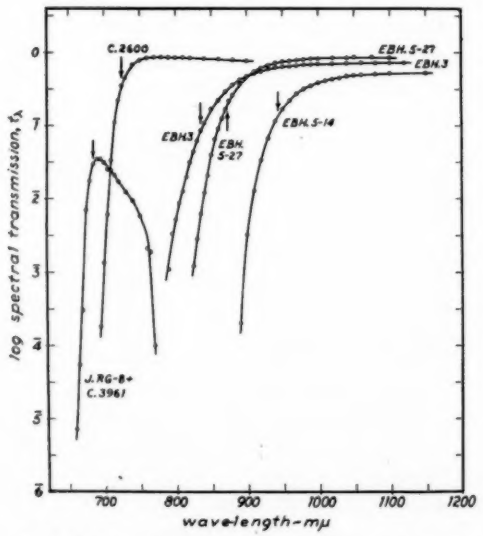


FIG. 2. Spectral radiant transmissions of five of the filters used in the measurement of visual thresholds in the red and infra-red. Arrows show the wave-lengths of maximum visual transmission for each filter.

To isolate the infra-red radiation for these measurements, we employed a graded series of infra-red filters with exceptionally sharp cut-offs, manufactured by the Polaroid Corporation (Blout *et al.*, 1946). In the present paper these filters are arbitrarily designated with the notation EBH and a number. They provided bands of radiation which were limited on the short wave-length side by the cut-off of the filter, and on the long wave-length side by the steeply falling sensitivity of the eye. They had maximum spectral visual transmissions, as finally determined, at 835, 873, 908, 928, and 944 mμ. A 685 mμ filter (Jena RG8 plus Corning 3961) was used as a standard reference red; since it transmitted a band from 660 to 750 mμ its visual transmission could be calculated from the familiar sensitivity function for the ordinary visible spectrum. The visual transmissions of the other filters were always computed relative to this red reference filter. To fill out the spectral range we also used a Corning 2600 filter, with maximum spectral visual transmission at 725 mμ. Figure 2 shows the spectral (radiant) transmissions of five of these filters, measured with a Beckman spectrophotometer.

In calculating the visual transmission of a filter, one multiplies together at convenient wave-length intervals the following three quantities:

(1) The spectral (radiant) transmission of the filter, t_λ,
(2) The relative spectral energy distribution of the source R_λ, and,
(3) The relative spectral sensitivity of the eye, K_λ.[2]

These are all pure numbers, the spectral transmission being the ratio of transmitted to incident radiation or illumination at each wave-length and the spectral energy emission and spectral sensitivity being expressed as fractions of the maximal values of those functions.

The product $R_\lambda t_\lambda K_\lambda$ expresses the relative luminous flux or relative visual effectiveness of the radiation transmitted at wave-length λ from

[2] The terms, *spectral luminosity* and *spectral sensitivity*, are used interchangeably to denote the reciprocal of the relative energy at each wave-length needed to produce a constant visual effect. The term *luminosity* directs attention primarily to the radiation, *sensitivity* to the properties of the eye.

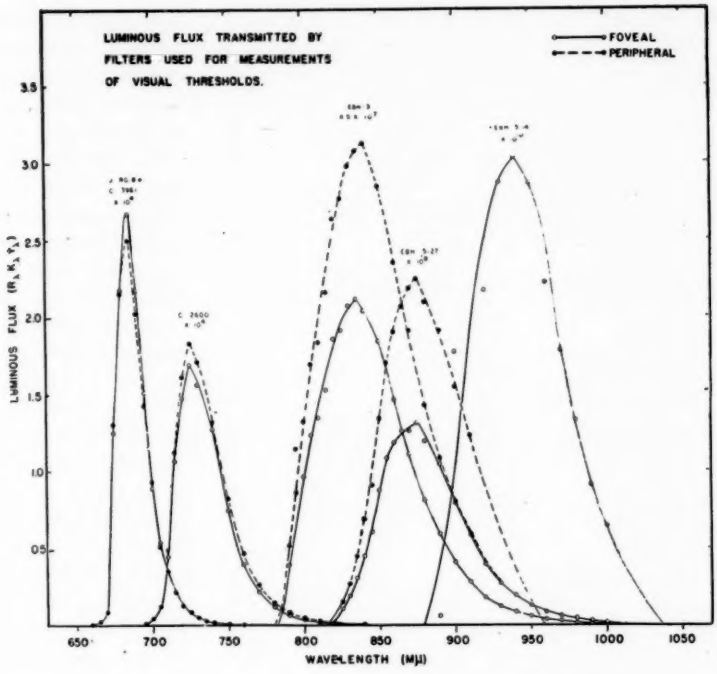


FIG. 3. Spectral visual transmissions of the five filters, the radiant transmissions of which are shown in Fig. 2. The visual transmissions are different for foveal (cone) and peripheral (rod) vision; and these differences increase toward the infra-red, as the peripheral retina grows more sensitive relative to the fovea. Source: blackbody radiation at 3200°K.

a given type of source. It is this product which we call the spectral visual transmission, to distinguish it from the total visual transmission defined below.[3] It also, of course, is a pure number. At shorter wave-lengths this product is reduced by the low values of t_λ; at long wave-lengths it is diminished by the rapidly decreasing values of K_λ.

In Fig. 3 are plotted the values of $R_\lambda t_\lambda K_\lambda$ for the same five filters shown in Fig. 2. (The values for K_λ used in Fig. 3 are those obtained at the end of the investigation.) Figure 3 shows the bands of luminous flux passed by each filter, and it is clear that most of the effective luminous flux is contained within a sharply peaked band 50 to 100 mμ in width.

The total luminous flux transmitted by each filter is given by $\int R_\lambda t_\lambda K_\lambda d\lambda$, while the flux incident on the filter, i.e., that characteristic of the unfiltered source, is given by $\int R_\lambda K_\lambda d\lambda$. The visual transmission of the filter, T_v, is the ratio of the transmitted to the incident flux, or:

$$T_v = \frac{\int R_\lambda t_\lambda K_\lambda d\lambda}{\int R_\lambda K_\lambda d\lambda}.$$

Thus one could compute T_v provided that he knew R_λ, t_λ, and the sensitivity function, K_λ. Our problem, however, was the converse one: given values of T_v from the threshold measurements, to find a sensitivity function yielding calculated values of T_v which agree with the measurements.

We first computed trial values of T_v for each filter, using the values of t_λ shown in Fig. 2, the appropriate values of R_λ from the radiant energy spectrum of a blackbody at 3200°K, and a sensitivity function derived from Goodeve's previous measurements in the infra-red (1936). These computed values of T_v were then compared with those obtained from the threshold measure-

[3] The total visual transmission or luminous transmission of a filter is given by the expression $T_v = \int R_\lambda t_\lambda K_\lambda d\lambda / \int R_\lambda K_\lambda d\lambda$. It would seem obvious, therefore, to define the visual transmission at each wave-length as $(R_\lambda t_\lambda K_\lambda / R_\lambda K_\lambda)$; but this would equal simply the radiant transmission, t_λ, which is, of course, the same whether measured by eye or with a thermopile (cf. Committee on Colorimetry, 1944, p. 256). Such a definition of spectral visual transmission would be useless. We have chosen instead the very useful expression for this term indicated above.

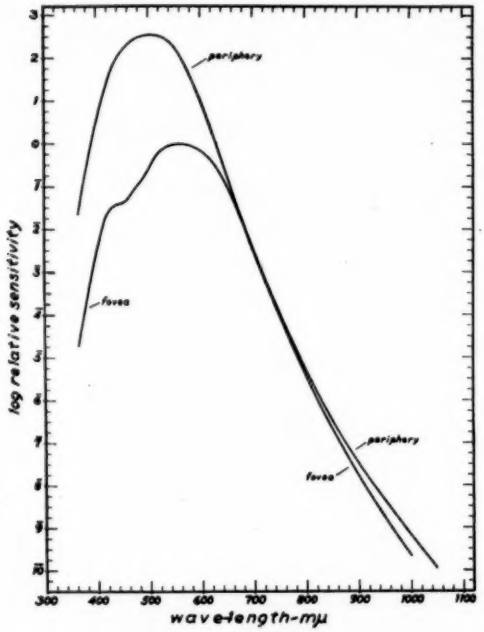


FIG. 4. Relative spectral sensitivity of the dark adapted fovea and peripheral retina. The spacing of these curves is appropriate for 1° test fields fixated within the fovea or 8° above the fovea, and exposed for 1 second. Below 750 mμ, the foveal curve is a composite function based on the original data of a number of workers; the peripheral function is from Wald (1945b). To these the data for the far red and infra-red have been joined.

ments. Corrections could then be applied to the sensitivity function, if necessary, until a function was obtained which produced agreement between observed and calculated values of T_v.

FOVEAL SENSITIVITY IN THE INFRA-RED

The measurements reported by Goodeve (1936) apparently were essentially foveal.[4] Two principal observers and an unspecified number of others were employed, and the visual threshold was measured by changing the wave-length, with the source constant. The image observed was the exit slit of a monochromator, a rectangle about forty times as high as it was wide and subtending a visual angle of 6.2°×0.15°. The eye was dark-adapted, fixation was purposely avoided, and the infra-red stimulus was reported to have various colors from red to yellow. Since Goodeve's sensitivity function appeared to have been obtained largely with foveal vision, we used it tentatively to calculate visual transmissions of our filters. In so doing we joined Goodeve's function to a preferred foveal sensitivity function for the visible spectrum.[5]

We also obtained measured values of T_v for several filters by means of visual threshold determinations, using a one degree circular field fixated foveally and exposed for one second. The average thresholds of 39 to 54 observers yielded values for the visual transmissions of filters which are shown in Table I. This table also shows calculated values of T_v for the same filters.

The differences between observed and calculated values are small and show no systematic trend. We may conclude that Goodeve's function is an adequate description of foveal sensitivity in the infra-red. Since the computation of the visual transmission of the densest filter (No. EBH 5–27) required values of K_λ at wave-lengths as long as 1000 mμ, it was necessary to extrapolate Goodeve's sensitivity function beyond 900 mμ. This offers no difficulties since it

TABLE I. Observed and calculated visual transmissions, foveal.

Observations with a 1-degree field, fixated within the fovea, and exposed for 1 second. Calculated visual transmissions, (T_v), are based on foveal values of relative sensitivity (K_λ) taken from Fig. 4, and on the energy distribution of a black body at 3200°K. λ_{max} is the wave-length of maximum spectral visual transmission of each filter (Fig. 3). In the designations of the filters, J = Jena, C = Corning, and EBH designates a special group of infra-red filters manufactured by the Polaroid Corporation.

Filter	JRG-8 plus C 3961	C 2600	EBH 3	EBH 5-27
λ_{max}	685	725	835	873
No. of observers	54	48	39	54
I. Calc. log T_v relative to 685 mμ filter	0.00	+0.03	−3.30	−3.90
II. Obs. log T_v relative to 685 mμ filter	0.00	−0.09	−3.23	−3.91
III. $\log \frac{\text{calculated}}{\text{observed}}$ values (I–II)	0.00	+0.12	−0.07	+0.01

[4] As noted above, this is true also of the work of Pinegin (1945).

[5] This foveal sensitivity curve, shown in Fig. 4, is a composite of the original measurements of Hyde and Forsythe (1915), Coblentz and Emerson (1918–19), Gibson and Tyndall (1923), Jainski (1938), Fedorov *et al.* (1940), and Wald (1945b). Its derivation will be described in detail elsewhere. The large depression and minor inflections in the region 440–520 mμ mark the absorption bands of xanthophyll, the pigment of the human yellow patch (Wald, 1945b).

is linear when plotted on a frequency basis. This extrapolated portion of the foveal function yielded a calculated visual transmission in good agreement with the measurements.

The complete foveal sensitivity function for the red and infra-red is shown in Fig. 4, joined to the foveal function for the ordinary visible spectrum discussed above. The values themselves are given in Table V.

PERIPHERAL SENSITIVITY IN THE INFRA-RED

When calculating the visual transmissions of filters for peripheral vision, we assumed as a starting point that the peripheral sensitivity function would parallel Goodeve's curve which had proved satisfactory for the fovea. Table II shows the results of such computations, together with the observed values of visual transmission of the same four filters listed in Table I. We can see at once that there are considerable discrepancies between observed and calculated values and that the sensitivity function must therefore be corrected.

The discrepancies in Table II show that the periphery was more sensitive than the assumed function had predicted, and this was also evident from a direct comparison of foveal and peripheral thresholds. Such a comparison is presented in Table III, which reveals a progressive increase in sensitivity of the periphery as compared with the fovea as one goes to longer wave-lengths.

TABLE II. Observed and calculated visual transmissions, peripheral.

Observations made with a 1-degree field, fixated 8 degrees above the fovea and exposed for one second. Source at 3200°K; 2 cm of water in light path. Otherwise as in Table I.

Filter	JRG-8 plus C 3961	C 2600	EBH 3	EBH 5-27
λ_{max}	685	725	838	874
No. of observers	54	48	39	54
I. Obs. log T_v relative to 685 mμ filter	0.00	+0.02	−3.07	−3.68
II. Calc. log T_v based on Goodeve's sensitivity function	0.00	+0.01	−3.27	−4.01
III. Discrepancy (I minus II)	0.00	+0.01	+0.20	+0.33
IV. Calc. log T_v based on corrected sensitivity function	0.00	+0.09	−3.10	−3.62
V. Discrepancy (I minus IV)	0.00	−0.07	+0.03	−0.06

TABLE III. Variation, with respect to wave-length, of the ratio between foveal and peripheral thresholds in the red and infra-red.

Measurements with the adaptometer described in text, using a 1-degree field and 1-second exposures. Peripheral fields fixated 8 degrees above the fovea.

Filter	JRG-8 plus C 3961	C 2600	C 2540 plus C 2600	EBH 3	EBH 5-27	EBH 5-36
λ_{max}	685	725	762	835	873	908
No. of observers	54	48	10	39	54	4
$\log \frac{\text{peripheral threshold}}{\text{foveal threshold}}$	+0.04	−0.08	−0.08	−0.16	−0.20	−0.34

The ratios between foveal and peripheral thresholds shown in Table III were used to correct the assumed peripheral function; the corrected sensitivity curve thus obtained yielded the calculated values of T_v listed in the fourth row of Table II. These corrected values are in satisfactory agreement with the measurements, the discrepancies being 0.07 log unit or smaller and showing no consistent trend.

We were able to extend the measurements of visual thresholds with the peripheral retina further into the infra-red than was possible with the fovea, both because of the greater sensitivity of the periphery and because larger retinal areas could be stimulated. In order to accomplish this we used a magnifying ocular between the observer's eye and the stimulus field so that we obtained a 7 degree field centered 8 degrees above the fovea. With this arrangement we could measure visual thresholds with filters having their maximum spectral visual transmissions at 908, 928, and 944 mμ; the results of these measurements are shown in Table IV.

In order to calculate the visual transmissions of these denser filters it was necessary to extend the corrected peripheral sensitivity function described above to 1050 mμ. The resulting function yielded calculated values of T_v which showed close agreement with the measurements. We feel justified, therefore, in presenting this function in Fig. 4 and in Table V as the best available description of the sensitivity of the peripheral retina from 700 to 1050 mμ. In Fig. 4 it has been joined to a peripheral rod sensitivity function for the ordinary visible spectrum (Wald, 1945b).

It should be noted in Fig. 4 that the foveal function is almost completely fused with the peripheral between 675 and 740 mμ. This may

TABLE IV. Observed and calculated visual transmissions, peripheral.

Data from 11 observers, using a 7-degree field, fixated peripherally and exposed for 1 second.

Filter	EBH 5–36	EBH 5–8	EBH 5–14
λ_{max}	908	928	944
I. Obs. log T_ν relative to 873 mμ filter	−0.75	−1.18	−1.44
II. Calc. log T_ν relative to 873 mμ filter, based on corrected sensitivity function	−0.76	−1.14	−1.46
Discrepancy (I minus II)	+0.01	−0.04	+0.02

appear to contradict a previous statement that in the far red the fovea is more sensitive than the periphery by about 0.2 log unit (cf. Wald 1945b). The difference between Fig. 4 and the earlier data lies in the exposure times employed. The earlier work was done with 1/25 second exposures, whereas the present measurements involved one second exposures, since we wished to obtain the lowest possible thresholds in order to push the determinations as far as possible into the infra-red.

In a special series of measurements designed to test this point we compared 1-second exposures with 1/25-second flashes using both the adaptometer described in this paper and also the instrument employing a mercury arc which was described earlier (Wald 1945a). With both instruments the foveal and peripheral thresholds at 685 or 691 mμ were within a few hundredths of a log unit of each other with 1 second exposures, while with 1/25 second exposures the fovea was more sensitive, on the average, by 0.25 log unit.

In addition to the effect of exposure time, one must also consider that with fields smaller than 1 degree the fovea is favored relative to the periphery. That is, in fields of this size the foveal threshold rises much more slowly than the peripheral threshold as the diameter of field is reduced below 1° (cf., for example, Graham and Bartlett, 1939). Consequently the greater sensitivity of the periphery in the infra-red may well disappear when either smaller test fields or shorter exposure times are used. This is probably why Goodeve and others who have studied the visibility of infra-red radiation have not reported the increasing sensitivity of the rods relative to the cones, as one goes to longer wave-lengths.

THE SPECTRAL LIMITS OF VISION

In closing, it is interesting to inquire whether there is any end to the process of extending the spectral range of human vision by employing more and more intense sources of radiation. One could probably make threshold determinations beyond 1050 mμ under the proper conditions,

TABLE V.

Spectral sensitivity (K_λ) in the red and infra-red. Values selected as described in the text, for 1° test fields, exposed for 1 second, and fixated either within the fovea or 8° above the fovea in the wholly dark-adapted eye. The foveal values beyond 750 mμ are identical with those of Goodeve (1936). All spectral sensitivities are stated relative to the maximum sensitivity of the fovea, here taken as 1 ($\log K_\lambda = 0.0$); on this basis the maximum sensitivity of the peripheral area, at 505 mμ, is 363 ($\log K_\lambda = 2.56$).

Wave-length (mμ)	1° foveal field		1° field fixated 8° above fovea	
	$\log K_\lambda$	K_λ	$\log K_\lambda$	K_λ
700	$\bar{3}.55$	3.55×10^{-3}	$\bar{3}.53$	3.4×10^{-3}
710	$\bar{3}.23$	1.7	$\bar{3}.23$	1.7
720	$\bar{4}.92$	8.3×10^{-4}	$\bar{4}.92$	8.3×10^{-4}
730	$\bar{4}.61$	4.1	$\bar{4}.63$	4.3
740	$\bar{4}.31$	2.0	$\bar{4}.34$	2.2
750	$\bar{4}.02$	1.04	$\bar{4}.06$	1.15
760	$\bar{5}.71$	5.1×10^{-5}	$\bar{5}.78$	6.0×10^{-5}
770	$\bar{5}.43$	2.7	$\bar{5}.51$	3.2
780	$\bar{5}.15$	1.4	$\bar{5}.25$	1.8
790	$\bar{6}.87$	7.4×10^{-6}	$\bar{6}.99$	9.8×10^{-6}
800	$\bar{6}.59$	3.9	$\bar{6}.73$	5.4
810	$\bar{6}.34$	2.2	$\bar{6}.48$	3.0
820	$\bar{6}.08$	1.2	$\bar{6}.23$	1.7
830	$\bar{7}.83$	6.8×10^{-7}	$\bar{7}.99$	9.8×10^{-7}
840	$\bar{7}.57$	3.75	$\bar{7}.76$	5.75
850	$\bar{7}.34$	2.2	$\bar{7}.53$	3.4
860	$\bar{7}.11$	1.3	$\bar{7}.32$	2.1
870	$\bar{8}.88$	7.5×10^{-8}	$\bar{7}.11$	1.3
880	$\bar{8}.65$	4.5	$\bar{8}.90$	7.9×10^{-8}
890	$\bar{8}.45$	2.8	$\bar{8}.71$	5.1
900	$\bar{8}.23$	1.7	$\bar{8}.52$	3.3
910	$\bar{8}.03$	1.07	$\bar{8}.35$	2.2
920	$\bar{9}.82$	6.6×10^{-9}	$\bar{8}.17$	1.5
930	$\bar{9}.62$	4.2	$\bar{8}.00$	1.0
940	$\bar{9}.43$	2.7	$\bar{9}.82$	6.6×10^{-9}
950	$\bar{9}.24$	1.7	$\bar{9}.65$	4.5
960	$\bar{9}.05$	1.1	$\bar{9}.48$	3.0
970	$\bar{10}.87$	7.4×10^{-10}	$\bar{9}.31$	2.0
980	$\bar{10}.69$	4.9	$\bar{9}.14$	1.4
990	$\bar{10}.51$	3.2	$\bar{10}.98$	9.55×10^{-11}
1000	$\bar{10}.34$	2.2	$\bar{10}.82$	6.6
1010			$\bar{10}.66$	4.6
1020			$\bar{10}.51$	3.2
1030			$\bar{10}.35$	2.2
1040			$\bar{10}.19$	1.55
1050			$\bar{10}.04$	1.1

but eventually one would reach a point beyond which the eye ceases to be a specialized sense organ which is more sensitive to radiant energy than the general body surface.

The human skin can respond to electromagnetic radiation, both visible and infra-red, by giving rise to the sensation of warmth, provided that the intensity of the radiation is sufficiently great. This is why infra-red radiation is often called radiant heat.[6] The most extensive data on the threshold of the skin to heat radiation are those of Hardy and Oppel (1937). They found that the threshold is inversely proportional to the area stimulated, over a wide range of areas. Their values for the threshold ranged from 0.019 g cal./cm²/sec. for 0.2 cm² of the skin of the forehead to 0.00015 g cal./cm²/sec. for the entire posterior body surface of a naked man.

It is probably unprofitable to attempt a direct comparison between the retina and the skin, since the density of receptor cells in the skin is low and variable, compared with the retina, and because the skin is not equipped with a refracting system to concentrate the radiation upon a discrete area. It is possible, however, to compare the eye as a whole with the general body surface, as represented by the undifferentiated skin. In this case we are interested in the energy falling on the cornea on the one hand and on the skin on the other.

Hecht, Shlaer, and Pirenne (1942) give the absolute threshold of human rod vision at 510 mμ under optimal conditions as about 3×10^{-10} ergs delivered in a 1/1000-second flash. This is equivalent to 1.85×10^{-14} g cal./cm²/sec. falling on the cornea. However, it is well known that the threshold intensity could have been reduced by a factor of about 100 if exposures of 1 second or longer were used instead of the 1/1000-second flash (Graham and Margaria, 1935). This would give a threshold for the rods at 510 mμ of 1.85×10^{-16} g cal./cm²/sec. From Fig. 4 we see that at 1050 mμ the retina is less sensitive than this by a factor of 3.3×10^{12} which indicates a threshold at this wave-length of 0.00061 g cal./cm²/sec. This is slightly higher than Hardy and Oppel's threshold for the entire posterior body surface of a man, but considerably lower than their values for small areas of skin. Figure 4 shows that in the infra-red the sensitivity of the eye falls by a factor of ten when the wave-length increases by about 60 mμ, so that by 1150 mμ one would expect the sensitivity of the eye to drop below that of even small areas of skin.

Another factor which operates in favor of the skin is the absorption spectrum of water itself. This is negligible at wave-lengths shorter than 950 mμ, but at 1150 mμ according to Hartridge and Hill (1915), all but 15.9 percent of the radiation entering the eye is absorbed by the water of the ocular media before reaching the retina, while at 1200 mμ only about 7.9 percent reaches the rods or cones. Thus it seems probable that the absorption of water, together with the rapidly falling sensitivity of the eye, sets a limit at about 1150 or 1200 mμ beyond which incident radiation is more readily felt as heat than seen as light.

[6] Thomas Young (1802) stated the basic relation between light and heat radiation very soon after the discovery of the infra-red by Herschel in 1800: ". . . the affections of heat may perhaps hereafter be rendered more intelligible to us; at present, it seems highly probable that light differs from heat only in the frequency of its undulations or vibrations; those undulations which are within certain limits, with respect to frequency, being capable of affecting the optic nerve, and constituting light; and those which are slower, and probably stronger, constituting heat only; . . . ." The present experiments extend into the curious region in which radiation may be regarded with equal justice as light or as radiant heat.

BIBLIOGRAPHY

W. de W. Abney and W. Watson, "The threshold of vision for different coloured lights," Phil. Trans. Roy. Soc. London, **A216**, 91 (1915).

E. R. Blout, W. F. Amon, Jr., R. G. Shepherd, Jr., A. Thomas, C. W. West and E. H. Land, "Near infra-red transmitting filters," J. Opt. Soc. Am. **36**, 460 (1946).

W. W. Coblentz and W. B. Emerson, "Relative sensibility of the average eye to light of different colors and some practical applications to radiation problems," Bull. Bur. Stand. **14**, 167 (1918–1919).

Committee on Colorimetry, "The psychophysics of color," J. Opt. Soc. Am. **34**, 245 (1944).

A. Dresler, "Beitrag zur Photometrie farbiger Lichtquellen, insbesondere zur Frage des Verlaufs der spektralen Hellempfindlichkeit," Das Licht **7**, 81–85 (1937).

N. T. Fedorov, V. I. Fedorova, A. G. Plakhov, and L. O. Seletzkaya, "A new determination of the relative luminosity curve," J. Physics Acad. Sci. U.S.S.R. **3**, 5 (1940).

K. S. Gibson and E. P. T. Tyndall, "Visibility of radiant energy," Sci. Pap. Bur. Stand. **19**, 131 (1923).

C. F. Goodeve, "Vision in the ultra-violet," Nature **134**, 416–417 (1934).

C. F. Goodeve, "Relative luminosity in the extreme red," Proc. Roy. Soc. London **A155**, 664–683 (1936).

C. H. Graham and N. R. Bartlett, "The relation of size of stimulus and intensity in the human eye: II. Intensity thresholds for red and violet light," J. Exper. Psychol. **24**, 574 (1939).

C. H. Graham and R. Margaria, "Area and the intensity-time relation in the peripheral retina," Am. J. Physiol. **113**, 299–305 (1935).

D. R. Griffin, R. Hubbard, and G. Wald, "The sensitivity of human rod and cone vision to infra-red radiation," J. Opt. Soc. Am. **36**, 360 (1946).
J. D. Hardy and T. W. Oppel, "Studies in temperature sensation. Part III. The sensitivity of the body to heat and the spatial summation of the end organ responses," J. Clin. Investigation **16**, 533–540 (1937).
H. Hartridge and A. V. Hill, "Transmission of infra-red by the media of the eye and the transmission of radiant energy by Crookes and other glasses," Proc. Roy. Soc. London **B89**, 58 (1915).
S. Hecht, S. Shlaer, and M. H. Pirenne, "Energy, quanta, and vision," J. Gen. Physiol. **25**, 819–840 (1942).
E. P. Hyde and W. E. Forsythe, "The visibility of radiation in the red end of the visible spectrum," Astrophys. J. **42**, 285 (1915).
P. Jainski, "Die Spektrale Hellempfindlichkeit des menschlichen Auges und ihre Bedeutung für die Lichtmesstechnik," Diss. Tech. Hochschule, Berlin, 1938.
N. I. Pinegin, "Absolute sensitivity of the eye in the infra-red spectrum," C. R. Acad. Sci. U.R.S.S. **47**, 627 (1945).
G. Wald, "The spectral sensitivity of the human eye. I. A spectral adaptometer," J. Opt. Soc. Am. **35**, 187–196 (1945a).
G. Wald, "Human vision and the spectrum," Science **101**, 653–658 (1945b).
T. Young, "On the theory of light and colours" (Bakerian Lecture), Phil. Trans. Roy. Soc. London **92**, Part I, 12 (1802).

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 37, NUMBER 7 JULY, 1947

The Photoelectric Measurement of Fluorescence Spectra

R. A. BURDETT AND L. C. JONES, JR.
Wood River Research Laboratories, Shell Oil Company, Inc., Wood River, Illinois

(Received March 29, 1947)

An attachment is described which makes possible the measurement of fluorescence spectra with the Beckman Spectrophotometer. The fluorescence spectra of a few aromatic hydrocarbons are presented.

INTRODUCTION

PHOTOELECTRIC measurements of emission and absorption spectra have recently become quite common, and even the weak Raman spectra have been measured by this means.[1] Zworykin[2] and Studer[3] have described equipment suitable for the measurement of cathodoluminescence spectra. However, no apparatus is generally available for the photoelectric measurement of fluorescence spectra excited by ultraviolet light. The authors have felt, therefore, that it might be of interest to describe a technique which has been used in these laboratories for some time.

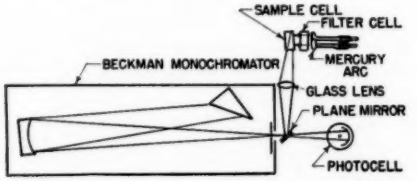


FIG. 1. Experimental arrangement for determination of fluorescence spectra.

The widely available Beckman Quartz Spectrophotometer[4] has been adapted to the determination of fluorescence spectra by means of a very simple attachment which can be interchanged with the usual light sources for absorption studies. Intensity measurements are made directly with the photo-cells ordinarily supplied with the instrument. No changes in the measuring

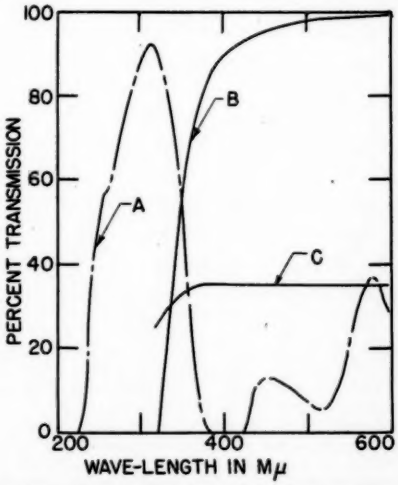


FIG. 2. Absorption spectra of parts of optical system: A. Ni, Co filter solution, B. lens, C. lamp envelope.

[1] D. H. Rank and R. V. Wiegand, J. Opt. Soc. Am. **36**, 325 (1946).
[2] V. W. Zworykin, J. Opt. Soc. Am. **29**, 84 (1939).
[3] F. J. Studer, paper presented before Opt. Soc. Am., February 21, 1947.
[4] H. H. Cary and A. O. Beckman, J. Opt. Soc. Am. **31**, 682–689 (1941).

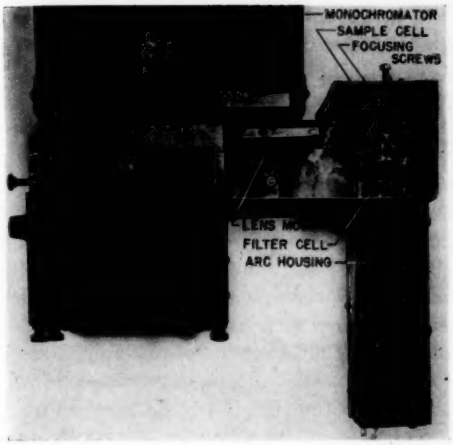


FIG. 3. The fluorescence adapter.

circuits are necessary as it is possible to determine the spectra of all of the condensed ring polynuclear aromatic hydrocarbons and of many other compounds of similar fluorescent power with the unmodified instrument. However, it is planned to install cells of the photo-multiplier type for use in the investigation of very weak fluorescences.

EXPERIMENTAL DETAILS

The experimental arrangement is shown in the diagram of Fig. 1. The fluorescence is excited by means of appropriately filtered mercury arcs, two of which have been used in this work. A low pressure quartz mercury arc in the form of a 2-turn flat spiral $1\frac{1}{4}''$ in diameter is used as a source of the Hg 2537A line.[5] It operates from a 3000V, 30 ma transformer. A liquid filter consisting of a 1 cm thickness of a solution of 200 grams of $NiSO_4 \cdot 6H_2O$ and 50 grams of $CoCl_2 \cdot 6H_2O$ per liter of 1:4 H_2SO_4 removes most of the visible and longer wave-length ultraviolet lines from the exciting radiation.[6] The absorption spectrum of this filter is shown in Fig. 2.

A Hanovia Sc 5030, Utility Model Quartz Lamp, filtered by Corning Glass Filter No. 5840 (red ultra), furnishes 3650A excitation. The liquid filter is removed when the high pressure arc is in use.

Both the filter and the sample cells are conventional 1.0-cm thick, 2.5-cm diameter fused quartz absorption cells closed by a ground glass stopper.[7] The usual arrangement of a horn-shaped fluorescence cell placed directly before the spectrometer slit has been discarded in favor of the scheme shown, which takes advantage of the observation of Lommel[8] that "the intrinsic intensity (of surface fluorescence) increases rapidly with increasing angle of observation."

In this investigation the concentration of the fluorescent material has been adjusted so that most of the emission occurs in a layer a millimeter or less thick at the front of the cell. The cell is tilted slightly as shown in Fig. 1, and an image of the fluorescent layer is formed at the entrance slit of the monochromator by a MgF coated glass achromatic lens of 49-mm focal length and 18-mm diameter. This lens limits the measurements to wave-lengths longer than 3200A.

All of the optical elements are mounted in a sheet steel box which is fastened to the monochromator by a locking plate mechanism identical with those ordinarily used to mount the visible and ultraviolet light sources for absorption studies. The adapter with cover and mercury lamp removed is shown in Fig. 3.

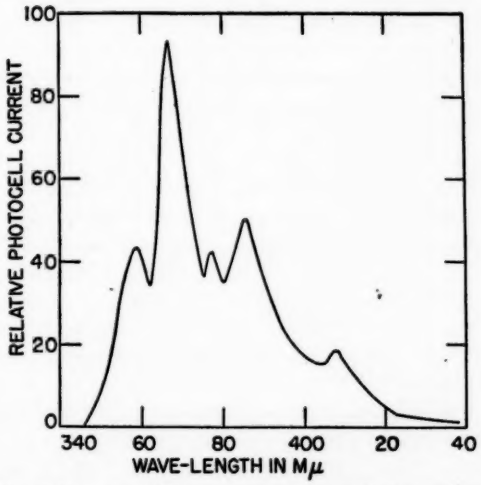


FIG. 4. Uncorrected fluorescence spectrum of phenanthrene in benzene solution, 1 mg/ml.

[5] This lamp was made to order by the Hanovia Chemical and Manufacturing Company, Newark, New Jersey.

[6] P. Pringsheim and M. Vogel, *Luminescence of Liquids and Solids* (Interscience Publishers, Inc., New York, 1943), p. 41.

[7] These cells are available from the American Instrument Company, Silver Spring, Maryland.

[8] F. Lommel, Wied. Ann. **10**, 449 (1880); cf. R. W. Wood, *Physical Optics* (The Macmillan Company, New York, 1934), third edition, p. 656.

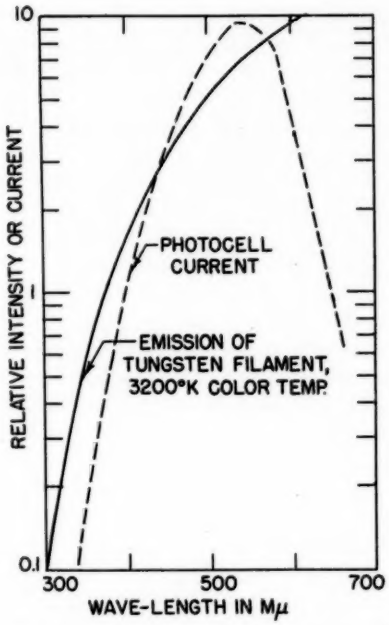


FIG. 5. Data for construction of intensity correction curve.

MEASUREMENTS AND CALCULATIONS

The fluorescent solution is introduced into the sample cell by means of a hypodermic syringe. The measuring circuit is then adjusted to maximum sensitivity and the range selector set at 0.1. The ultraviolet sensitive photo-cell has been used for most of these measurements.

Fixed slit-widths of 0.1–0.3 mm, corresponding to half-intensity band widths of 0.5–1.5 mμ at 400 mμ, are usually employed. With these slit-widths none of the Hg lines from the filtered low-pressure arc can be detected when the cell is filled with a non-fluorescent liquid. With wider slits, however, one can detect apparent fluorescence bands in the blank spectrum which are in reality caused by scattered radiation from the mercury arc.

The relative photo-cell current corresponding to each wave-length is read directly from the transmission scale of the instrument. For most practical purposes a direct plot of these readings against wave-length is adequate. Such a plot for a solution of phenanthrene in benzene is given in Fig. 4. If, however, the true energy distribution curve is desired, correction must be made for the effects of variations in photo-cell sensitivity and in the linear dispersion of the monochromator with wave-length. Loss of light by absorption and reflection in the lens must also be considered.

A curve has been derived which gives the total correction for all of these effects as a function of wave-length. It was obtained in the following manner: a tungsten lamp, of 3200K color temperature, whose spectral energy distribution could be calculated, was placed before the entrance slit of the monochromator, and the resulting photo-cell current was measured as a function of wave-length at a fixed mechanical slit-width. The absorption spectra of a section of the lamp envelope and of the fluorimeter lens were measured with the spectrophotometer. These data are shown in Figs. 2 and 5. The correction curve was calculated from the expression

$$F_\lambda = E_\lambda T_{g\lambda} / I_\lambda T_{l\lambda}$$

where F_λ is the factor to convert photo-cell current, I_λ, to relative intensity of the tungsten filament, E_λ, $T_{g\lambda}$ is the transmittancy of the glass lens, and T_λ the transmittancy of the lamp envelope, all at the wave-length λ.

Figure 5 gives the energy distribution for a tungsten filament at 3200K color temperature. The data for wave-lengths greater than 350 mμ

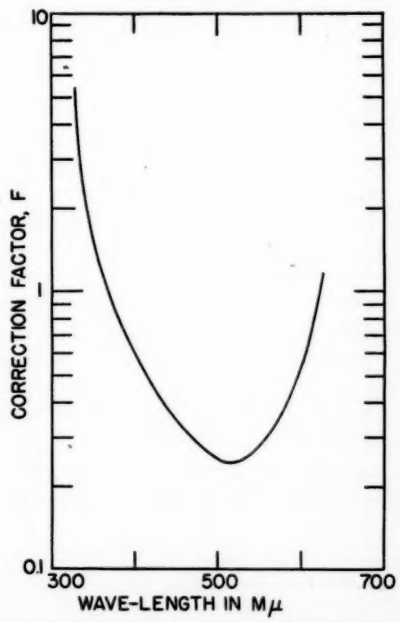


FIG. 6. Curve for conversion of photo-cell current to relative intensity.

re from the work of Forsythe and Adams.[9] An identical curve extended to 320 mμ was obtained by computation from the published data on the conversion of color to absolute temperature[10] and on the relative emissivity of tungsten filaments and of a blackbody radiator.[11,12]

The energy distribution for the blackbody at 3100K (the absolute temperature of a tungsten filament at 3200K color temperature) was obtained by extrapolation of the data at lower temperatures by means of Wien's formula for isochromes.[13]

$$\log E_\lambda = C_1 - C_2/\lambda T$$

where C_1 and C_2 are constants and T is the absolute temperature.

The final correction curve is presented in Fig. 6, and selected values for F are given in Table I.

TABLE I. Intensity correction factors for blue sensitive photo-cell.

Wave-length in mμ	Correction factor F	Wave-length in mμ	Correction factor F
320	12.0	500	0.255
340	2.39	520	0.240
360	1.19	540	0.260
380	0.800	560	0.305
400	0.610	580	0.385
420	0.470	600	0.525
440	0.380		
460	0.315		
480	0.275		

TABLE II. The wave-lengths of the fluorescence bands of aromatic hydrocarbons.

Hydrocarbon	In mineral oil solutions This investigation	In mineral oil solutions Dragulescu	In benzene solutions this investigation
Naphthalene	338.5 mμ		
	343		
	349		
Phenanthrene	359		358.5
	366		367
	377		377
	385		385
	408		408
Anthracene	403	403	405
	426	426	428
	452	453	454
	484	481	487

[9] W. E. Forsythe and E. Q. Adams, J. Opt. Soc. Am. **35**, 108–113 (1945).

[10] *International Critical Tables* (McGraw-Hill Book Company, Inc., New York 1929), Vol. 5, p. 246.

[11] See reference 10, pp. 240–241.

[12] F. Hoffman and H. Willenberg, Physik Zeits. **35**, 713–15 (1934).

[13] C. Schaefer and F. Matossi, *Das Ultrarote Spektrum* (J. Springer, Berlin, 1930), p. 91.

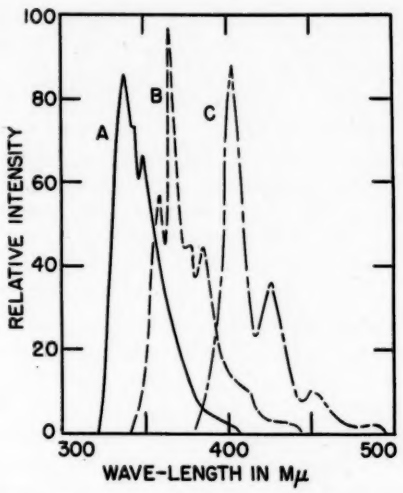


FIG. 7. Fluorescence spectra of several aromatic hydrocarbons: A. naphthalene, B. phenanthrene, C. anthracene.

FLUORESCENCE SPECTRA

The corrected fluorescence spectra of aromatic hydrocarbons given in Fig. 7 are typical examples of the sort of information obtainable with this apparatus. The spectra shown are those of 0.1 mg/ml solutions of naphthalene, phenanthene, and anthracene in a highly refined mineral oil. The low pressure arc was used to excite the fluorescence. Use of the 3650A source gave an identical spectrum for anthracene but no measurable fluorescence for naphthalene or phenanthrene since they do not absorb in this region of the spectrum.

The wave-lengths of the fluorescence bands found along with those reported by Radulescu and Dragulescu[14] are summarized in Table II. It will be seen that the bands found are in good agreement with those reported in the older investigation.

The hydrocarbons were specially purified for this investigation. Their melting points and ultraviolet spectra indicated a high degree of purity.

It will be noted that the fluorescence spectra are highly characteristic and show much less mutual interference than the corresponding absorption spectra. This suggests that such data may be of considerable value in qualitative identifications.

[14] D. Radulescu and C. Dragulescu, Bull. Soc. Chim. Rom. **17**, 9–37 (1935).

Improved Infra-Red Polarizer[1]

A. H. PFUND
Johns Hopkins University, Baltimore, Maryland
(Received March 22, 1947)

The original selenium polarizer has been found to suffer from two defects: (1) the degree of polarization is not sufficiently high, and (2) the necessary dimensions become unwieldy. Both defects are overcome by resorting to two reflections from parallel selenium mirrors. The efficacy of this procedure is proved by a study of the dichroism of beryl. A discussion of selenium mirrors produced by vacuum distillation follows.

THE selenium polarizer was first introduced by the writer in 1906[2] when it was shown that, at the same polarizing angle, the device was effective from the visible spectrum out to about 12μ. Since then, Rubens[3] and his co-workers have extended this work to much greater wave-lengths. In the meantime two defects in the original design have become apparent: first, the degree of polarization is not sufficiently high; secondly, in parallel light, obtained from a concave mirror 10 in diameter, the selenium reflector reaches the unwieldy dimensions of 10×30 cm. Both of these defects were overcome by resorting to the use of a double-mirror polarizer. While this modification involves a moderate loss of intensity, the gain is distinctly worth while.

The basic argument underlying these considerations may perhaps best be introduced by a discussion of the Fresnel curves for the intensities of reflection of the two polarized components for selenium. These curves, shown in Fig. 1 were calculated for the refractive index $n=2.90$ and for the constant polarizing angle of 71°.

Since in practice the polarization is not complete even at the polarizing angle, calculations were made for the intensities resulting from reflection from two parallel selenium mirrors. The corresponding ordinates, marked $(R_{\parallel})/(R_{\perp})^2$ are introduced to the right of the diagram. Assuming that a depolarization of 0.01 is permissible, it is evident from this last curve that the permissible degree of depolarization is not exceeded in the angular range of incidence 59° to 77°. This means that convergent instead of parallel light may now be used, in consequence of which the dimensions of the polarizers may be reduced greatly.

The modified mounting, based on the above considerations, is shown in Fig. 2 where S is an incandescent source whose radiations, after reflection from the silvered concave mirror M_1, are brought to a focus between the two parallel selenium polarizers P_1 and P_2. The silvered concave mirror M_2 finally brings the rays to a focus at F. As before[2] the selenium mirrors were prepared by the method of casting—but with this significant difference: a selenium mirror several millimeters thick loses its figure with time, whereas a much thinner mirror does not. The latter, however, will be so transparent to infra-red radiations that some energy will be reflected from the selenium-glass interface. Around 9μ, where glass has a band of selective reflection, the radiation will be depolarized markedly. To overcome this difficulty, the glass is first rough-ground with coarse carborundum before applying the selenium.

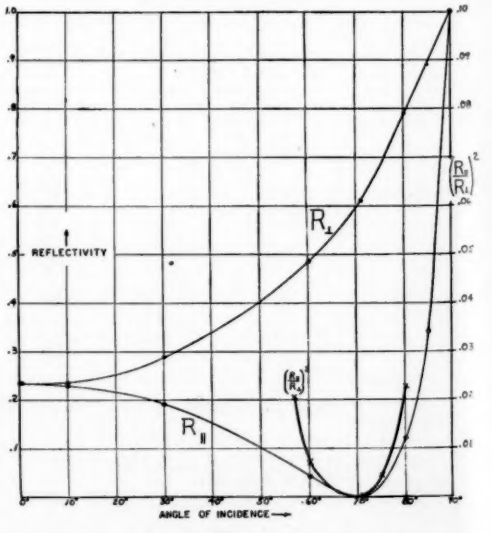


FIG. 1.

[1] A. H. Pfund, J. Opt. Soc. Am. **36**, 715(A) (1946).
[2] A. H. Pfund, Ap. Jn'l. **XXIV**, 19 (1906).
[3] H. Rubens, see Schaefer and Matossi, *Das Ultrarote Spektrum*, Chapter 14 (1930).

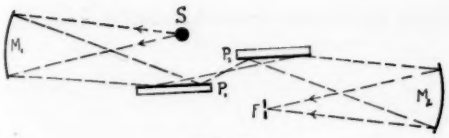


FIG. 2.

This procedure, which might be termed "internal roughening," introduces the large retardation $\Delta = 2nt\cos r$* between the disturbances reflected from troughs and crests.[4] As a result of this procedure, regular reflection at the selenium-glass boundary is eliminated—at least up to and beyond 14μ.—After preparation of the rough glass surface, molten selenium is poured on the (hot) ground surface and a larger and thinner (heated) glass plate is placed on top. Vigorous pressure is applied so that the thickness of the selenium layer is reduced to a fraction of a millimeter. Care must be exercised in not allowing the selenium and the glass plates to get too hot since, under such conditions, the selenium, on cooling, is liable to assume the allotropic (metallic) state which is undesirable. After the system has cooled down to room temperature, the upper plate is bent gently, thus causing it to let go of the selenium.

Tests on the effectiveness of this new double-mirror polarizer were carried out on a clear crystal of beryl (3 $BeO \cdot Al_2O_3 \cdot 6\ SiO_2$) having an effective thickness of 12 mm. As shown by Koenigsberger,[5] this uniaxial crystal reveals marked dichroism in the region 0.7 to 1.7μ. For this study a glass (Fuess) monochromator together with such accessories as tungsten lamp, vacuum thermopile and galvanometer was employed. For all determinations of "transmission" the crystal was mounted in front of the entrance slit of the monochromator. The first test involved a determination of the transmission curves (for the ordinary and extraordinary rays) upon polarizing the light with a large Nicol prism. Measurements were then repeated upon replacing the Nicol by the selenium polarizer. The results are presented in Fig. 3 where the continuous curve applies to the Nicol prism determinations whereas the dots represent the results obtained with the selenium polarizer. The conclusion to be drawn is that, within the limits of accuracy, the two sets of measurement are identical and that the degree of polarization achieved with the selenium is as good as that realized with the Nicol prism.

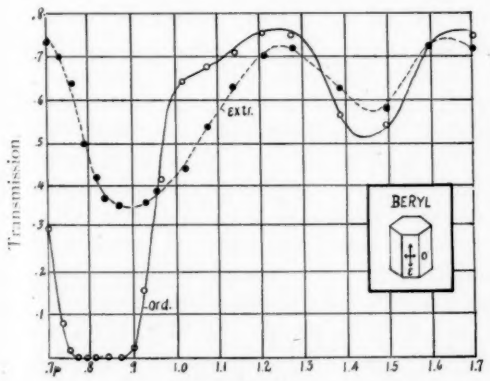


FIG. 3.

It has long been known that the method of vacuum evaporation yields mirrors of almost unlimited size and of superb quality. Consequently, layers of selenium, deep red in transmitted light, were distilled on flat surfaces of glass and of lucite. In both cases the results were bad since the reflection from the lower surface of the selenium produced interference bands (with low intensities in the minima) and, in the case of glass, also produced pronounced selective reflection (depolarized) near 9μ. To overcome these difficulties it was suggested by my colleague, Dr. John Strong, that two tungsten spirals filled, respectively, with selenium and with some (non-reactive) metal, be mounted under the same bell-jar. To begin with, much metal and little selenium would be distilled simultaneously. The metal would, presumably, come down in the "black" condition[6] whence the radiation would be absorbed and not reflected. The proportion of metal would be decreased gradually so that the upper surface would contain only selenium. Unfortunately this suggestion has not been tried out as yet. As some time must elapse before this work can be taken up again, it seemed desirable to put on record the results thus far obtained.

* Here t = height of elevations of ground surface, n = refractive index of selenium ($n = 2.9$), r = angle of refraction within selenium.

[4] Lord Rayleigh, "Polish," Phil. Mag.; see also R. W. Wood, Phys. Optics, second edition, p. 45.

[5] J. Koenigsberger, Ann. d. Physik **61**, 687 (1897).

[6] A. H. Pfund, Rev. Sci. Inst. **1**, 397 (1930). A. H. Pfund, J. Opt. Soc. Am. **23**, 375 (1933).

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 37, NUMBER 7 JULY, 1947

Spectral Variations of the Intermittency Effect for Some Panchromatic Plates

W. HOPWOOD AND J. D. CRAGGS
High Voltage Laboratory, Research Department, Metropolitan-Vickers Electrical Company, Ltd., Manchester, England
(Received March 25, 1947)

Experiments have been made to determine the response of photographic panchromatic emulsions to light flashes lasting only a few microseconds. Previous experiences of other workers and general photographic theory tended to show that the spectral response of plates should show no intermittency effect, i.e., should be the same for short flashes of light and for steady illumination. This was found not to be so and experimental data are provided.

I. INTRODUCTION

THE response of photographic plates to short flashes of light is of technical interest and importance in a number of ways. For example, much work has been carried out during the last 5 years, in British and American laboratories, on the behavior of intense light sources operated for very short times, and the behavior of photographic emulsions in such circumstances has had to be investigated. However, the present experiments were undertaken in connection with spectroscopic studies of hydrogen spark channels using discharges which last only for 1–20 microseconds. It is normal, in spectroscopic analyses of the usual kind, to obtain plate calibrations in terms of reference lines produced by the specimen under test and so to avoid the necessity for finding the frequency response of the plates used in relation to the time of duration of the discharges (sparks or arcs). In fact, spectral calibrations of plates for such very short flashes of light do not appear to have been published.

In the present case, measurements were undertaken on the intensity distribution in the Balmer series, for reasons to be discussed in another publication by the present authors. The work therefore involved a knowledge of the spectral response of the plates to very short light pulses, since the most convenient manner of calibrating the plates was to refer them to an NPL calibrated tungsten filament source running at a known temperature. It was clearly preferable to expose the plates to the tungsten lamp for a single, short, continuous period on the assumption that the plates would show no spectral variations of reciprocity failure and intermittency effect. Berg[1] has recently shown that the reciprocity failure vanishes for very short light flashes, i.e., so short that ionic movements, cf. electronic movements, are negligible. Measurements were made with light flashes lasting from $\frac{1}{3}$ to 1/300,000 sec., but no data on spectral variations were reported. The importance of the work is that it shows that the behavior of emulsions, at least so far as reciprocity failure is concerned, is known in general for times of the order of those involved in our experiments.

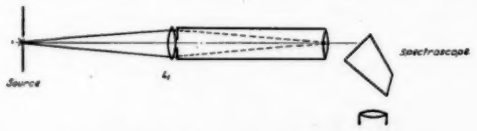


FIG. 1. Optical system used in plate tests.

It is also known that reciprocity failure shows no spectral response, so far as the existing measurements are concerned. Webb[2] gives curves that demonstrate this fact clearly for the plates he used and for a range of times from 1000 sec. to 1/30 second. The $\log It/\log I$ and $\log It/\log t$ curves for equal blackening are of the same shape at different wave-lengths. This shows that, on the basis of equal blackening, a single short flash and a single long exposure will give the same spectral variations of energy. Since it is necessary in practice to expose the plates for perhaps 1000 flashes, it is also necessary to show that the spectral responses for one and for a superposition of say 1000 flashes are the same. The latter problem appears to have received much less attention than the former, although both are clearly inter-connected. However, Webb[2] shows that, again on a basis of equal density, curves

[1] W. F. Berg, Proc. Roy. Soc. **A174**, 559 (1940).
[2] J. H. Webb, J. Opt. Soc. Am. **23**, 316 (1933); **25**, 4 (1935).

connecting logIt and flash frequency (1:4 light-to-total period) for different wave-lengths also are of the same shape. The upper limit of flash frequency is not given, but is stated to be beyond the critical frequency at which the reciprocity failure disappears. It is also difficult to assess the accuracy of Webb's results, since the spectral responses are plotted on a coarse scale. For example $[\log It(\lambda=6910)/\log It(\lambda=5460)]=1.6/0.6$, i.e., ~ 2.5 for a flash frequency >1/sec. In our experiments it was felt that errors of not more than ± 10 percent should be permitted, and for this reason it was felt that the performance of a separate series of experiments was justified, although great encouragement was derived from Webb's work.

Other papers having a bearing on the work are cited by Webb[2] and in the recent work of Mees.[3]

2. ARRANGEMENT OF APPARATUS

The optical system is shown in Fig. 1 and exhibits little of special interest.

An image of the source, whether spark or tungsten filament, was cast onto the collimator lens by the condenser which was placed immediately before the spectrograph slit. This system insures uniform illumination of the slit, and because the light passing through the collimator is condensed, avoids reflections inside the collimator tube. The spectrograph was a small Hilger constant deviation instrument, with a camera attachment.

The hydrogen discharge exhibited some unusual characteristics. Its duration was from 1 to 20 microsec. controlled by the circuit adjustments, and a square current wave was utilized. Details of the circuit, triggering gap, and of the discharges are given in other papers.[4–7]

3. DETAILS OF THE EXPERIMENTS

Photographs of the spectrum of the hydrogen discharge, and of the continuous spectrum from a tungsten filament lamp running at a known temperature were taken on the same photographic plate. The continuous spectrum was intended for calibration purposes since the distribution of energy in the spectrum at a given temperature, in this case 2700°K, is known. Three spectra were taken on each plate, one each of the discharge and of the continuum both taken through a Hilger neutral step wedge which was mounted

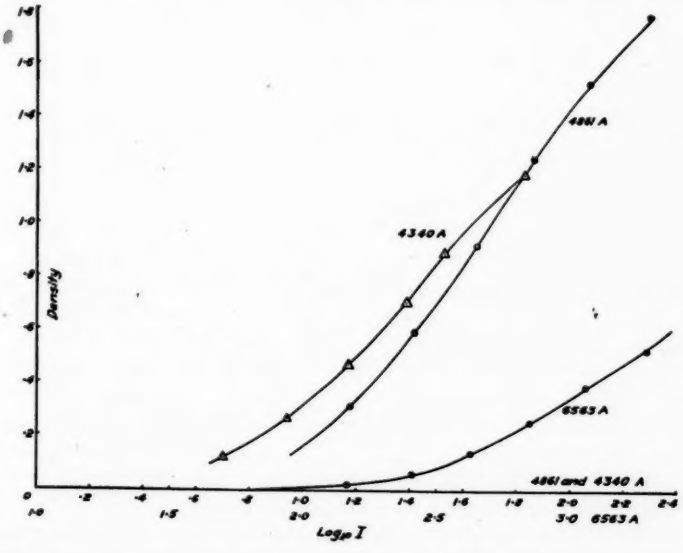


FIG. 2. Plate response with tungsten lamp. All three curves are derived from one plate (one exposure) and correction is made for varying emission with wave-length (from published data on emission characteristics of tungsten filaments).

[2] C. E. K. Mees, *The Theory of the Photographic Process* (The Macmillan Company, New York, 1944).
[4] J. D. Craggs and J. M. Meek, Nature **156**, 21 (1945).
[5] J. D. Craggs and J. M. Meek, Proc. Roy. Soc. **A186**, 241 (1946).
[6] J. D. Craggs, M. E. Haine, and J. M. Meek, J. Inst. Elec. Eng. (to be published).
[7] J. D. Craggs and W. Hopwood (to be published).

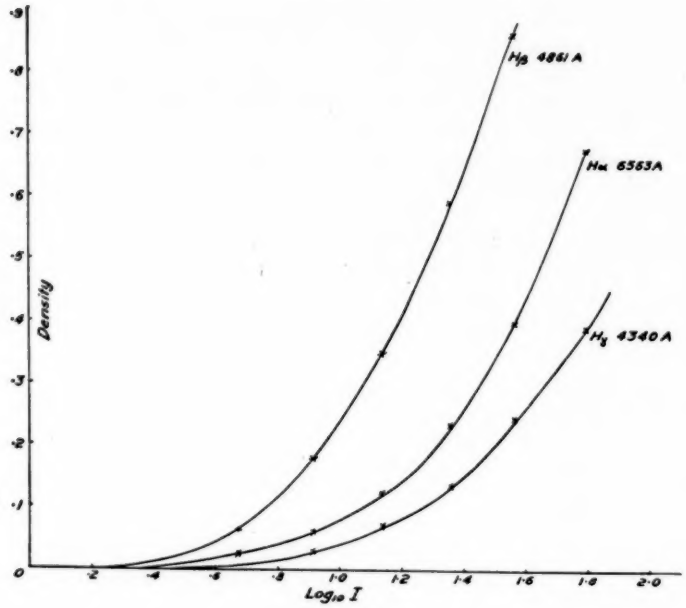


FIG. 3. Plate response with hydrogen sparks. (1-microsec. duration, run at a recurrence rate of ~100 sec.)

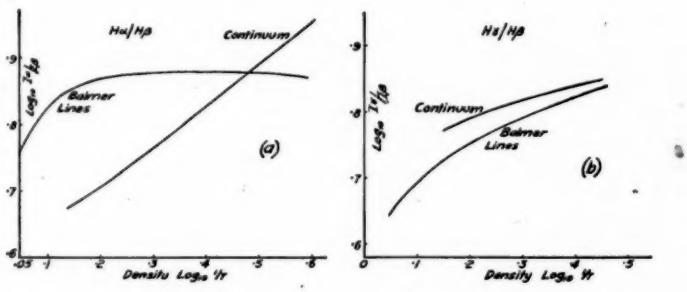


FIG. 4. Differential spectral response of plates to intermittent and steady light. (4-microsec. pulses.)

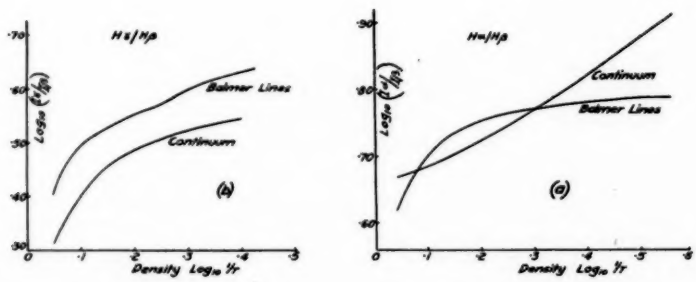


FIG. 5. Differential spectral response of plates to intermittent and steady light. (1-microsec. pulses.)

directly before the spectrograph slit, and a third one of the discharge taken without the step wedge in position, in order to check the uniformity of illumination along the length of the slit.

The photographic plates used in the experiments were Ilford Soft Gradation Panchromatic plates, which were chosen because their spectral response was such that they gave fairly equal densities for the Balmer lines in the spectrum of

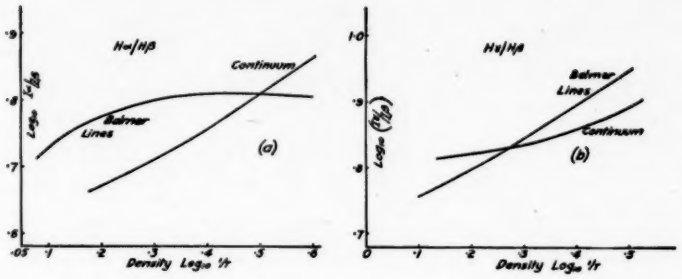


FIG. 6. Differential spectral response of a photographic plate to intermittent and steady light. (10-microsec. pulses.)

TABLE I.

	Density	Log10 I Hα	Log10 I Hβ	Log10 I Hγ	I Hα	I Hβ	I Hγ	I scaled to Hα Hα	I scaled to Hα Hβ	I scaled to Hα Hγ
Hydrogen spark illumination	0.1	1.08	0.76	1.26	12.0	5.75	18.2	1	0.48	1.52
	0.2	1.32	0.94	1.50	20.9	8.71	31.6	1	0.42	1.51
	0.3	1.46	1.07	1.66	28.8	11.7	45.7	1	0.41	1.58
	0.4	1.57	1.19	1.81	37.2	15.5	64.6	1	0.42	1.74
								λ=6563A	4861A	4340A
Tungsten filament lamp (continuous illumination)	0.1	2.55	0.94	0.65	355	8.71	4.47	1	0.0245	0.013
	0.2	2.75	1.06	0.84	562	11.5	6.92	1	0.0205	0.012
	0.3	3.08	1.26	1.09	1200	18.2	12.3	1	0.015	0.010
	0.4	3.38	1.42	1.30	2400	26.3	20.0	1	0.011	0.0083

the discharge. The plates were developed in a normal M.Q. developer and were lightly brushed with a pad of cotton wool during development to insure even development.

Density measurements were carried out on the plates on a Hilger non-recording microphotometer at the University of Manchester. Recorded profiles of the Balmer lines, which were broadened by the Stark effect,[7] were taken on a recording camera fitted to the instrument.

Calibration curves were plotted for the photographic emulsion at the Balmer wave-lengths 6563, 4861, and 4340A, for both the discharge spectrum and the continuum. The curves were plotted with density $[\log_{10}(I/T)]$ as ordinates and $\log_{10}P$ as abcissae, P being the percentage transmission of the wedge step. The step wedge had six steps of percentage transmissions 62.5, 37.1, 22.8, 13.7, 8.2, 4.7 percent, this enabling six points to be taken for each curve. Typical curves are shown in Figs. 2 and 3.

TABLE II.

Density	$\frac{I\text{H}\alpha}{I6563\text{A}}$	$\frac{I\text{H}\beta}{I4861\text{A}}$	$\frac{I\text{H}\gamma}{I4340\text{A}}$
0.1	1	1.9_5	117
0.2	1	2.0_5	126
0.3	1	2.7_4	158
0.4	1	3.8_2	210

Now if as is widely assumed there is no spectral variation in intermittency effect, the ratio of intensities at the three Balmer wave-lengths required to produce a given density should be the same for both series of curves. As the intensity distribution in the discharge spectrum is not known, a direct confirmation of the validity of this assumption cannot be made, but indirect evidence of its incorrect nature may be found in the following manner.

If any pair of wave-lengths is taken, e.g., 6563A and 4861A, and the difference ($\log_{10}P6563-\log_{10}P4861$) measured at some particular value of density, this expression will be equal to $[\log_{10}(I6503/I4861)-A]$, or to the logarithm of the ratios of the intensities required to produce that particular intensity minus a factor A which should be the same at all densities for that pair of wave-lengths, and is dependent only on the intensity distribution in the spectrum. Thus if curves are plotted for $\log_{10}(P\lambda1/P\lambda2)$ against density for both the Balmer lines and the continuum, and there is no spectral variation in

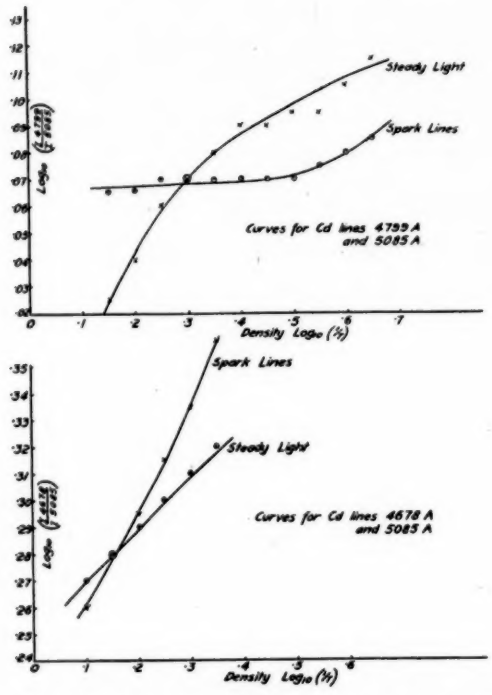

Steady Light
Spark Lines
Curves for Cd lines 4799 A and 5085 A
Density Log10 (1/T)
Spark Lines
Steady Light
Curves for Cd lines 4678 A and 5085 A
Density Log10 (1/T)

FIG. 7. Differential spectral response of a photographic plate to intermittent and steady light. (10-microsec. pulses.)

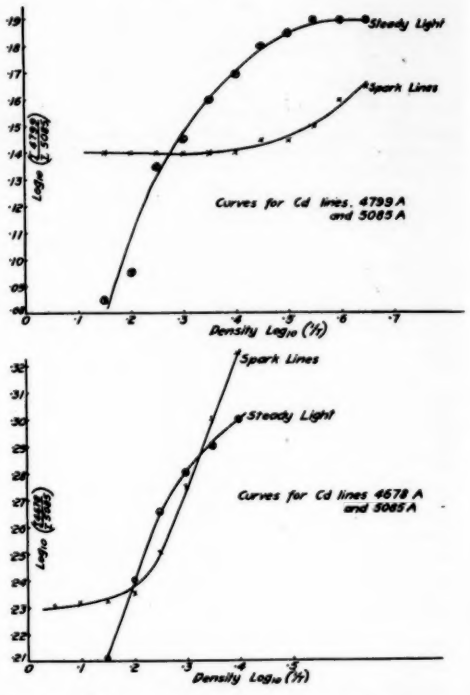

Steady Light
Spark Lines
Curves for Cd lines, 4799 A and 5085 A
Density Log10 (1/T)
Spark Lines
Steady Light
Curves for Cd lines 4678 A and 5085 A
Density Log10 (1/T)

FIG. 8. Differential spectral response of a photographic plate (different sample from that analyzed in Fig. 7) to intermittent and steady light. (10-microsec. pulses.)

the intermittency effect, these two curves should have the same shape, being only separated by a value of $\log_{10}(P\lambda 1/P\lambda 2)$ which is the same for all densities.

Such curves are shown in Figs. 4, 5, and 6 for three plates and using light pulses of varying duration and are plotted for the two pairs of lines, *viz.* 6563 and 4861A and 4340 and 4861A.

As can be seen the curves show quite large deviations from similarity in shape and show variations in separation as big as .15 in $\log_{10}(P\lambda 1/P\lambda 2)$, which corresponds to variations of 40 percent in the values of $P\lambda 1/P\lambda 2$, over the range of densities taken.

The data, plotted in Figs. 4–6, may be presented in an alternative manner, as shown in Table I and Table II for a representative plate.

The plates analysed in Figs. 4–6 were repeated many times with some months having elapsed between the series of experiments. The apparatus was dismantled and re-assembled several times, and many plates were studied. The effects shown in the figures were accurately reproducible, and it was therefore concluded that spectral decrements covering a wide range of wave-length as in the case of the Balmer series for hydrogen, could not be easily measured by photographic means for the very short discharges ($\sim$ 1 microsec.) in question.

Confirmatory results have been obtained in a series of experiments on the excitation of metallic electrode vapor in spark discharges of the type used here.[4–7] The results are plotted in Figs. 7 and 8 for the Cd lines 4799, 5085, and 4678A, for two different plates. Again, many confirmatory records were taken, and the effects shown are consistent from plate to plate, with variations of the order of those shown in Figs. 7 and 8. The plates used were Ilford HP3 Panchromatics.

It may be of interest to record that the spectral decrements for the short sparks have been measured in this laboratory with a photoelectric electron multiplier tube since 1943.[4,5] However,

lines having only slightly differing wave-lengths should still be accurately comparable with the photographic method since the spectral variation of plate response will then be small, and its variation with plate density will be of second order magnitude. Experiments on these lines are proceeding. Moreover, it is possible that other grades of plate may not show the effect described above. Ilford soft gradation plates were specially chosen in the Balmer series decrement experiments for their spectral response which falls rapidly at wave-lengths slightly less than 6563A and so enables good microphotometer records (i.e. of comparable density) of $H\alpha$ and $H\beta$ to be taken. This rapidly falling response may be responsible for the spectral variations of intermittency effect which may therefore not occur with plates of longer spectral response (such as Ilford Astra III or the excellent Ilford long range spectrographic emulsions).

ACKNOWLEDGMENTS

The authors are grateful to Mr. F. R. Perry for his interest in, and encouragement of the work. They also wish to thank Sir Arthur P. M. Fleming, C.B.E., D.Eng., Director of Research and Education, and Mr. B. G. Churcher, M.Sc., M.I.E.E., Manager of Research Department, Metropolitan-Vickers Electrical Company Ltd., for permission to publish this paper.

Exhibition of Scientific and Technical Photographs

A special section devoted to photographs of technical and scientific subjects will be included in the Annual Exhibition of Photography of the Photographic Society of America.

The exhibition will be hung at the Oklahoma Art Center, Oklahoma City, and contributors are invited to submit prints for the Technical Section prior to September 8. Prints should be mailed to W. F. Swann, 343 State Street, Rochester 4, New York.

As in past years, subject matter for the Technical Section may cover any phase of technical photography save pictorial photographs of technical or mechanical operations. Scientific and industrial photographs illustrating original and novel photographic techniques may be submitted. These include photographs in such fields as astronomy, geology, medicine (except diagnostic radiographs), metallography, mineralogy, and physics. Both black-and-white or color photographs are acceptable.

Technical data pertaining to each photograph should be lettered neatly in the lower left-hand corner of each mount, if possible, and the maker's name should appear on the face of the mount. There is no limit to the number of prints which may be submitted by a competitor.

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 37, NUMBER 7 JULY, 1947

Fluorescence of Silicate Phosphors*

KEITH H. BUTLER
Sylvania Electric Products, Inc., Salem, Massachusetts

(Received March 20, 1947)

An hypothesis is proposed for the fluorescence of silicate phosphors connecting the absorption and emission bands with the energy levels of the activator ions. It is shown that the energy levels of the ions must be radically changed as the ions enter the lattice of the solid.

In the case of Mn activated zinc silicate, the absorption of 2537A light is attributed to a transition from a $3d^5$ level to a $3d^44p$ level with an energy change of 40,000 cm^{-1} in the solid as opposed to 112,000 cm^{-1} for the free ion. The emission of visible light is due to a transition from an excited $3d^44p$ level to a $3d^44s$ level, after which there are non-radiating transitions to the unexcited level. Fine structure considerations indicate the possibility of two or more emission bands for excited manganese.

The results of beryllium additions to zinc silicate are discussed, and it is pointed out that the effectiveness in inducing new emission bands will vary with the cube root of the volume concentration. The behavior of alkaline earth silicates, activated by manganese, is also briefly discussed, and a reason for their excitation by cathode rays, but not by 2537 radiation, is suggested.

In lead activated alkaline earth silicates the light absorption is attributed to a transition from the $6s^2$ level to the $6s6p$ ($^1P_1°$) level, after which there is a non-radiating transition to the $6s6p$ ($^3P_0°$) level from which the electron falls back to the $6s^2$ level with radiation of blue or ultraviolet light.

I. INTRODUCTION

THE zinc sulfide and calcium sulfide phosphors, which are excited by long wave ultraviolet light, contain activators in such small amounts that they may be considered strictly as impurities. In contrast, the silicate phosphors, commonly used in fluorescent lamps, generally contain large amounts of activators, ranging from 1 to 5 percent by weight. In previous discussion on the theory of the fluorescence of silicate phosphors, it has been usual to assume the impurity picture developed by Seitz and others[1–4] for zinc sulfide phosphors. This impurity theory is unsatisfactory for the silicate phosphors for the following reasons:

1. It does not relate the color of the fluorescence to the particular activator used.
2. It does not serve to explain the effect of modifying atoms on the color of the fluorescence; for example, the effect of beryllium on zinc silicate activated by manganese.
3. It fails to explain why the same activator atom in various matrices is excited by different types of radiation; for example, manganese in zinc silicate is excited by both 3650 and 2537, while in calcium silicate it is excited only by cathode rays unless modifying atoms are present.
4. It fails also to explain the difference in color with different types of excitation; for example, calcium silicate activated by a combination of lead and manganese fluoresces yellow with cathode-ray excitation and orange-red with 2537.

If we return to fundamentals, a picture can be developed which explains some of these phenomena and may form the basis for the development of a quantitative theory by incorporation of the modern theory of solids.

Since, in general, an activator must be present for fluorescence to take place, it is obvious that the energy bands contributed by the activator determine the fluorescence and, since a particular activator gives a pre-determined color of fluorescence within a relatively small range of wavelengths, it is probable that the energy bands contributed by the activator are closely related to the energy levels of the free ions.

II. ENERGY RELATIONSHIPS OF IONS IN SOLIDS

Consider the effect of forming a solid on the energy relations, taking zinc oxide as a simple example. The energy of the valence electrons of a double ionized zinc atom in free space may be set as $E_1=-A$, where A is the quantized potential energy due to the attractive forces between the nucleus and the electron. In the solid this becomes $E_2=-A+R$, where R is a term due to the repulsive force between the valence electron and the surrounding negative field of the

* Presented at the meeting of the Optical Society of America in New York, New York, February 21, 1947.

[1] F. Seitz and R. P. Johnson, J. App. Phys. **8**, 84, 186, 246 (1937).
[2] N. F. Mott, Trans. Faraday Soc. **34**, 822 (1938).
[3] R. P. Johnson, J. Opt. Soc. Am. **29**, 387 (1939).
[4] R. C. Herman and C. F. Meyer, J. App. Phys. **17**, 743 (1946).

anions. For an excited electron these equations become $E_3 = -B$ and $E_4 = -B + nR$; here n is greater than one, since the repulsive force on the electron increases because of the increase in the radius of the orbit when the electron is excited, with deeper penetration into the negative field. The differences in energy terms, which govern the frequency of emission or absorption, are, respectively, $E_1 - E_3 = B - A$; $E_2 - E_4 = B - A - R(n-1)$. Thus the effect of placing the ion in the solid is to decrease the difference in energy between the normal and the excited state, and, as will be seen, this decrease must be of the same magnitude as the actual term value. Another effect is that the discrete levels become energy bands of some width since the attractive and repulsive forces vary over a range because of the close spacing of the ions, and if the levels are degenerate, they may split into several bands.

To estimate the relative magnitude of the attractive and repulsive forces on the valence electrons consider a zinc oxide molecule with one valence electron of the zinc ion lying on a straight line between the two nuclei. Its distance from the zinc nucleus approximates the ionic radius of Zn^{++} or 0.88A, while its distance from the oxygen nucleus approximates the ionic radius of 0^{--} or 1.76A. The attractive force is $-1 \times 1/0.88^2$, while the repulsive force is $+2 \times 1/1.76^2$, or about one-half the attractive force. So the assumption that the energy terms are materially changed between the free ion and the ion in the solid can be justified.

Considering in more detail the energy level picture, one of the basic selection rules accepted in quantum mechanics is that the electron must move from an "even" to an "odd" level or *vice versa*. The base level of most cations is "even," so the transition to an "odd" band must match the energy of the incident radiation in order for absorption to take place. If fluorescence is to take place, there must be an intermediate "even" band above the base state, and the transition to this state from the excited band is responsible for the fluorescent light. A slightly different picture holds for lead activation, as will be seen later.

The initial absorption of energy from the incident light may result from:

(a) The cations of the matrix.
(b) The cations of modifying ions included in the matrix.
(c) The activator ions themselves.

We will find examples of these three cases in common fluorescent materials:

(a) Zinc silicate, manganese activated.
(b) Calcium silicate, activated by manganese and lead.
(c) Barium silicate, activated by lead.

If the activator atom itself is the only absorber it will give up its energy promptly, and a rapid decay will result. However, if the base lattice or the modifying atom is the primary absorber the energy may travel in slow exciton waves, and the final transition will be delayed, resulting in a slow decay of fluorescence. Examples of the first type include unactivated zinc orthosilicate; calcium silicate, manganese activated; magnesium silicate, manganese activated; and barium silicate, lead activated. Examples of the second type include zinc silicate, manganese activated; and calcium silicate, activated by lead and manganese.

There may be certain instances in which no activator atoms are required for fluorescence, provided that the energy levels of the matrix cations conform to the requirements of absorption and emission, as outlined above. The tungstate phosphors may fall in this class though no attempt has been made to develop a theory for this type of phosphor.

III. ZINC ORTHOSILICATE, MANGANESE ACTIVATED

Consider now the specific case of manganese activated zinc silicate. An explanation of the absorption and emission bands can be found by consideration of the energy states of beryllium, manganese, and zinc.[5-9] The base state of beryllium III is an even level, and the next higher state is odd. The difference in energy of these is 1,240,000 cm^{-1} which would mean an absorption band for beryllium at an extremely short wavelength in the ultraviolet, consequently it cannot enter into the picture. In the case of zinc III, the base state is even. The first excited state, $3d^9 4s$, is even and has an energy of 80,000 cm^{-1} above the

[5] Bacher and Goudsmit, *Atomic Energy States* (McGraw-Hill Book Company, New York, 1931).
[6] K. C. Mazumder, Ind. J. Phys. **10**, 171 (1936).
[7] H. A. Robinson, Phys. Rev. **51**, 14 (1937).
[8] P. G. Kruger and H. T. Gilroy, Phys. Rev. **48**, 720 (1935).
[9] H. T. Gilroy, Phys. Rev. **38**, 2217 (1931).

base state. The first odd state, $3d^9 4p$, has an average energy of 141,000 cm^{-1} above the base state. For manganese III the base state is $3d^5$ and the next 2 higher states, $3d^4 4s$ (even) and $3d^4 4p$ (odd), have average energies of 64,000 and 112,000 cm^{-1} above the base state. Suppose we assume that because of the repulsive effect of the negative field surrounding the zinc ions in the solid, the $3d^9 4p$ state of zinc moves down to an energy level of about 55,000 cm^{-1}, then it will account for an absorption band in the ultraviolet at about 1800A units. When zinc silicate is excited by cathode rays or short ultraviolet light, the absorbed photon or electron raises an electron from the base state to the $3d^9 4p$ state, or a higher state of the zinc ion. This energy may remain localized at the excited ion or it may move through the lattice as an exciton wave. Eventually it becomes localized at a particular zinc ion, causing the formation of an excited state of the valence electron. The electron in this excited state will drop back to the $3d^9 4s$ level emitting energy (see Fig. 1). This will correspond roughly to an emission in the ultraviolet or short blue, and this is borne out by experimental evidence in that zinc silicate free of manganese can be feebly excited by cathode rays to emit blue light.

When we have a zinc silicate containing manganese, it is probable that the electrons in the base state of the manganese will be excited to the $3d^4 4p$ state, thus causing an absorption band at about 40,000 reciprocal centimeters or 2500A. This difference between the absorption of zinc silicate with and without manganese is borne out by experimental evidence from Kroger.[10] These

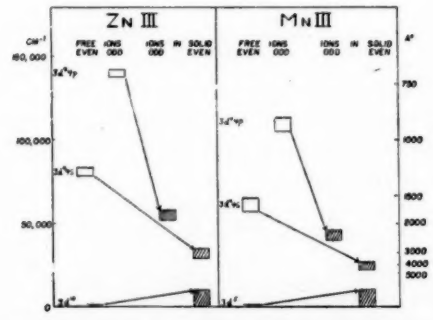


FIG. 1. Energy levels of zinc and manganese ions, free and in solid.

[10] F. A. Kroger, Physica **6**, 764 (1939).

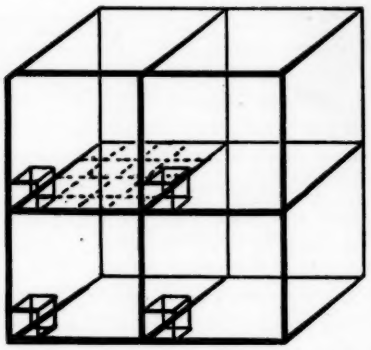


FIG. 2. Schematic lattice containing manganese ions.

excited Mn electrons will drop back from the $3d^4 4p$ state to the $3d^4 4s$ state, an allowed transition, and will emit light as a green band. From this $3d^4 4s$ state they will drop back to the base state, $3d^5$, by transitions which do not result in emission of visible light.

Some of the levels of manganese are degenerate, especially those of the $3d^5$ state. Such degenerate levels will split into several bands in the solid state. At the actual spacing of atoms in the solid some of these bands will merge into a single band while others will remain discrete. It seems probable that the lower levels, $3d^5$ and $3d^4 4s$, split in such a way that the transition from the excited to the unexcited state takes place in easy steps without radiation of visible light. For reasons which will appear later, it seems best to assume that the $3d^4 4s$ levels form discrete bands differing slightly in energy.

IV. EFFECT OF MODIFYING ATOMS

An elementary geometric consideration will show clearly the effect of the relatively high concentration of the activator and modifying ions. Suppose we consider that a lattice, Fig. 2, is composed of elementary building blocks and that these building blocks consist of a cubic array of zinc ions—4 on each side—together with their associated anions. Now if we replace a zinc ion at one corner of this elementary block by a manganese ion, the volume concentration of manganese is one atom in 64 or approximately 1.5 percent. However, when these blocks are assembled to form a lattice, the spacing between manganese ions along the x axis is one in four, so

that for a given direction through the lattice there must be a doubly periodic field of force; the shorter period being that of the action of the zinc nuclei on the electrons on which is superimposed a field of longer periodicity caused by the manganese nuclei. It is obvious that the linear concentration of the manganese ions in a given lattice will be equal to the cube root of the volume concentration; and starting with extremely low volume concentrations the linear concentration will rise rapidly and then flatten out in the well known form of a cubic curve.

Applying this relatively simple picture of the phosphor structure to zinc silicate, more of the phenomena experimentally found can be explained. Manganese and zinc, as they are found in this phosphor, are the same ionic diameter and are both doubly charged ions, consequently they will create approximately the same field of force in the lattice, and therefore introduction of additional manganese atoms will have relatively slight effect on the properties of the phosphor. This is in agreement with the experimental behavior of the zinc orthosilicate phosphor which is

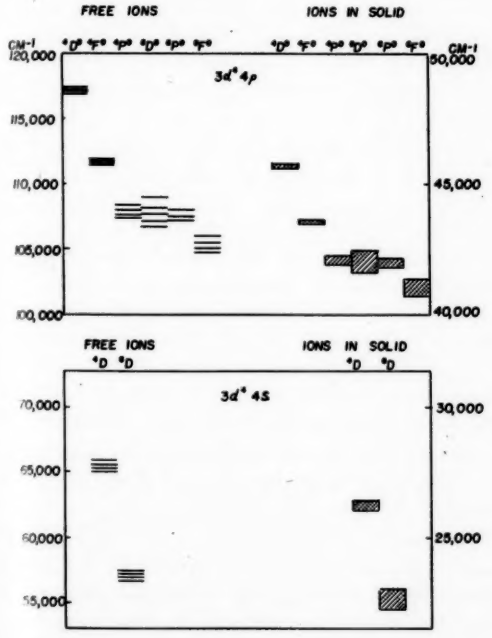


FIG. 3. Fine structure of energy levels of manganese ions, free and in solid.

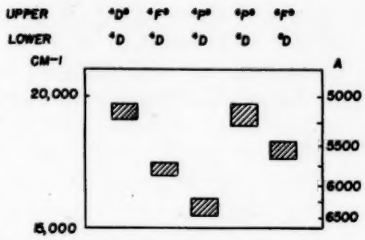


FIG. 4. Energy changes in electronic transitions for manganese ions in solid.

invariably green in color, but changes slightly toward the yellow with increasing Mn[10].

When we consider, however, the substitution of beryllium for zinc, it must be remembered that beryllium, although doubly charged, is a much smaller ion than zinc and will consequently change the field of force acting on the oxygen ions. This will in turn affect the interaction energy between the oxygen and the manganese, and thus the energy bands of the manganese will differ from those occurring in the beryllium free lattice. The splitting of degenerate levels may also be affected and the energy bands changed for this reason. However, these effects will follow the relation based on the cube root of the volume concentration, as outlined above; consequently a very small amount of beryllium will change the properties of the phosphor considerably, and with increasing amounts of beryllium the shift will become more and more gradual. Thus for zinc beryllium silicate, manganese activated, the introduction of beryllium should cause a shift in the absorption band of the phosphor toward longer wave-length, since the difference between the base state and the excited states of both Zn and Mn should be still further decreased by the influence of the strong secondary periodic field caused by beryllium. The introduction of beryllium should also shift the emission of manganese free zinc silicate. This shift in levels caused by beryllium will also be reflected in emission bands, when manganese is present, and there is ample experimental evidence for this shift.

With certain values for beryllium and manganese concentrations, it is found that a double peak[11] is formed in the emission spectrum. This can best be explained by the assumption of the

[11] J. H. Schulman, J. App. Phys. 17, 902 (1946).

existence of several discrete bands in the excited states, which can be excited by 2537 under suitable conditions, but with changing concentrations the degree of excitation varies, so that in emission various peaks are found. The reason for these several bands can be readily seen by referring to the fine structure of the manganese energy levels in Figs. 3 and 4, which are adapted from data by Gilroy.[9] The lowest state of MnIII is $^6S_{2\frac{1}{2}}$ so that the electron would normally move to one of the $^6P^\circ$ states, and if the transition were limited in this way, an emission band at about 5200A would be obtained. The fact that several bands are possible in zinc silicate with beryllium added is a good indication that the strong fields introduced by the beryllium make possible transitions to $6D^\circ$, $4D^\circ$ states, etc., which are not normally allowed.

The upper levels of excited manganese above the $3d^44p$ level are rather complex and as a result light of shorter wave-length than 2537 can be readily absorbed with non-radiating transition to the $3d^44p$ band. Thus zinc silicate, manganese activated, can be efficiently excited by rare gas resonance radiation and by cathode rays.

The theory presented above predicts fluorescence of zinc silicates without any added activator, assuming that the proper conditions of excitation are provided. In the case of zinc silicate, cathode rays provide this excitation, although the question of impurity activation is not unequivocably settled.

The theory of impurity levels previously proposed in the literature does find an application in explaining the strong phosphorescence induced by the introduction of arsenic into zinc beryllium silicate. Here the arsenic ions take interstitial positions in the lattice and introduce trapping centers at which the electrons are held until released by thermal or possibly infra-red excitation, lifting the electron to the excited state of the activator atom again.

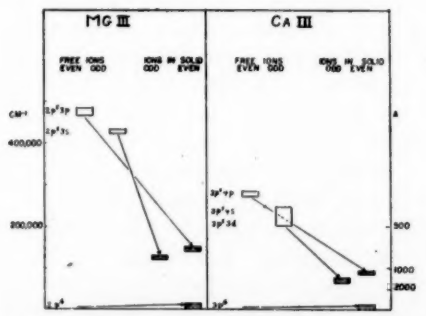


FIG. 5. Energy levels of magnesium and calcium ions, free and in solid.

To explain the fluorescence of zinc silicate, manganese activated, when excited by 3650 radiation, it may be assumed that some of the manganese atoms are so situated that they can absorb this energy and make the "forbidden" transition from $3d^5$ to $3d^44s$. They then make a further transition from $3d^44s$ to an upper level in the $3d^44p$ band by further absorption of 3650 radiation, after which they drop back to the bottom of this band and make a radiating transition to $3d^44s$ band. The whole process is relatively inefficient so that zinc silicates are poorly fluorescent to 3650 activation; also the activator atom acts as the absorber, and thus fluorescence decay is rapid.

V. ALKALINE EARTH SILICATES, MANGANESE ACTIVATED

The alkaline earth silicates, manganese activated, differ from zinc silicate in that they are activated by cathode rays with calcium silicate fluorescing yellow and magnesium silicate fluorescing red. The fluorescence decays more rapidly than is the case for zinc silicate. If we draw an energy level diagram, Fig. 5, for magnesium and calcium,[5] as modified for the solid state, it is obvious that the energy transitions are such that the matrix cations cannot absorb 2537 radiation. It may also be noted that because of the larger ionic radius of the alkaline earths the lattice of these silicates is considerably different from zinc orthosilicate, and the interaction effects on manganese will be different. We may assume that the energy bands of manganese will be shifted so that the $3d^94p$ band cannot absorb 2537 directly. The mechanism of excitation by cathode rays is then: (1) absorption of the cathode-ray electron by the matrix, (2) formation of a high energy exciton, (3) excitation of the manganese to a high energy band, (4) non-radiating transitions to the upper of the $3d^44p$ bands, and (5) radiating transition to the $3d^44s$ band with emission of yellow light.

If lead is introduced as a modifying agent into calcium silicate, manganese activated, the phos-

phor can be excited by 2537 with emission of red light. It is somewhat difficult to form a satisfactory picture to explain this phenomenon, but it is possible that lead introduces a secondary field sufficient to change the manganese bands, so as to allow absorption of 2537. It seems certain however that lead is the primary absorber when it is present in the lattice, since the phosphor containing lead has a slow decay rate.

VI. ALKALINE EARTH SILICATES, LEAD ACTIVATED

When we come to the lead activated alkaline earth silicate, a considerably different energy level diagram is found for lead than for manganese. In lead[8,12] the base state, $6s^2$ is even, while the first excited levels, $6s6p$, are odd as shown in Fig. 6. In the solid state these excited levels form three energy bands as indicated. When a photon of 2537 light is adsorbed it excites a $6s$ electron to the upper of the $6s6p$ bands. This excited electron then drops back to the lowest excited band without emitting visible light, and from this lower band it returns to the unexcited state with emission of long ultraviolet or of blue light. Because of the absorption of energy directly by the activator atom, no processes of long time constant are involved, and the fluorescence decays rapidly after removal of the exciting radiation.

Here again we can find evidence for the effect of modifying ions and of crystal structure. With the same activator concentration, calcium meta-silicate fluoresces in the ultraviolet and barium meta-silicate fluoresces blue. The emission of barium meta-silicate can be shifted toward the green by increase in the lead content. If, instead of the meta-silicate, the di-silicate $BaSi_2O_5$ is used as the matrix we find the emission takes place in the ultraviolet.

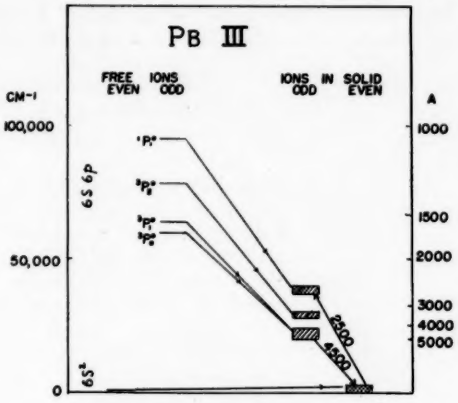


FIG. 6. Energy levels of lead ions, free and in solid.

Because of the rather simple energy level pattern of lead, it is not efficiently excited by either rare gas radiation or by cathode rays, since electrons excited to higher levels have no easy path by which they can return to the $6s6p$ band.

VII. CONCLUSION

A hypothesis relating the fluorescence of silicate phosphors to energy levels of the activator ions has been presented which gives reasonable explanations of a number of phenomena not satisfactorily explained by the impurity picture. Certain points developed in the discussion of this new picture are subject to experimental check and work, under way along this line, will be published in the near future.

[12] M. F. Crawford, H. B. McLog, and A. M. Crooker, Phys. Rev. **54**, 313(A) (1938).

Theoretical Resolution Angles for Point and Line Test Objects in the Presence of a Luminous Background

HOWARD S. COLEMAN AND MADELINE F. COLEMAN
The Optical Inspection Laboratory, The Pennsylvania State College, State College, Pennsylvania
(Received March 24, 1947)

A theoretical study has been made of the diffraction patterns for a circular aperture for point and line test objects in the presence of a luminous background. The study involves the superposition of the diffraction pattern elements of the test object and the image of the luminous background. From this superposition it is possible to relate the contrast of the image of the test object to the angle subtended by the test object elements and their apparent contrast at the entrance pupil of the optical system under consideration. This relation has been found useful in studying the resolving power of optical systems. For example, it is possible to determine the contrast of the image of a test object at the limit of resolution under a wide variety of conditions. It is also possible to determine the loss in resolution caused by a given reduction in image contrast regardless of the cause.

INTRODUCTION

IN experimental studies of the physical factors upon which the resolving power of optical systems depends, it is desirable to have a procedure of theoretically relating the minimum separable angle of the test object elements to the apparent contrast that a given test object may have with respect to its luminous background. It is the purpose of this paper to describe a method of obtaining such a relation for "point" and "line" test objects for optical systems having circular apertures, and for test objects having elements of equal brightness. The procedure can be extended to include apertures of other shapes and test objects having elements of unequal brightness.

THE PROCEDURE

The procedure consists of simply superposing the brightnesses of the diffraction patterns of the elements of the test object with the brightness of the image of the luminous background at the surface of the photosensitive receptor, used with the optical system under consideration. This can be visualized as the projecting of a test object on an illuminated screen, thus forming an image of the test object and the screen by means of the optical system under consideration. The superposition can be accomplished graphically as shown in Fig. 1. A series of such graphs can be made in which the angular separation of the test object elements and the background brightness is varied. These graphs can then be used to determine the brightness contrast of the superposed diffraction patterns and the image of the background for the particular conditions of interest.

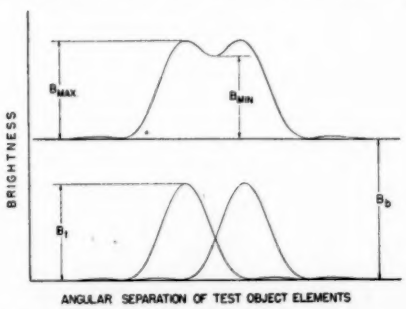


FIG. 1. Superposed diffraction patterns in the presence of a luminous background.

In order to execute the procedure outlined above, it is necessary to utilize a number of quantities which must be clearly defined. These are:

1. Coefficient of specific resolution.
2. Apparent contrast.
3. Test object contrast.
4. Diffraction pattern contrast.
5. Image contrast.
6. Apparent brightness of the test object.
7. Apparent brightness of the background.
8. Maximum brightness of superposed diffraction patterns.
9. Minimum brightness of superposed diffraction patterns.
10. The brightness of the image of the background.
11. The brightness of the diffraction pattern of one element of the test object.

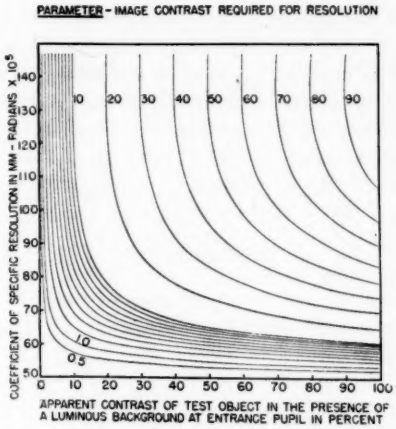

PARAMETER - IMAGE CONTRAST REQUIRED FOR RESOLUTION
COEFFICIENT OF SPECIFIC RESOLUTION IN MM - RADIANS X 10⁵
10 20 30 40 50 60 70 80 90
1.0
0.5
APPARENT CONTRAST OF TEST OBJECT IN THE PRESENCE OF A LUMINOUS BACKGROUND AT ENTRANCE PUPIL IN PERCENT

FIG. 2. Coefficient of specific resolution for a circular aperture *vs.* apparent contrast for point test objects in the presence of a luminous background.

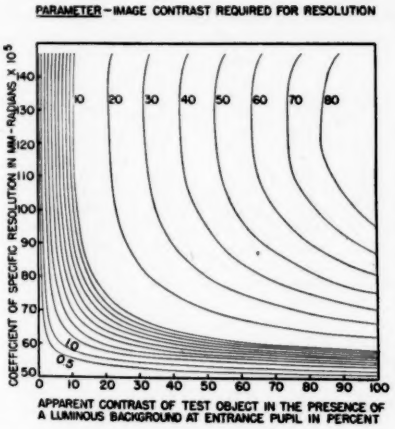

PARAMETER - IMAGE CONTRAST REQUIRED FOR RESOLUTION
COEFFICIENT OF SPECIFIC RESOLUTION IN MM - RADIANS X 10⁵
10 20 30 40 50 60 70 80
1.0
0.5
APPARENT CONTRAST OF TEST OBJECT IN THE PRESENCE OF A LUMINOUS BACKGROUND AT ENTRANCE PUPIL IN PERCENT

FIG. 3. Coefficient of specific resolution for a circular aperture *vs.* apparent contrast for line test objects in the presence of a luminous background.

The coefficient of specific resolution is defined as the product of the aperture, A, of the optical system under consideration and its minimum separable angle, α. This quantity has been assigned the symbol, $\bar{R}$, as indicated in Eq. (1), and is expressed in millimeters and radians.

$$\bar{R} = A\alpha. \tag{1}$$

These units have been found particularly convenient and are easily converted to other common units (e.g. inch-seconds).

The apparent contrast of the test object, C_a, is defined in terms of the apparent brightness, of the test object, b_a, at the entrance pupil of the optical system under test and the apparent brightness of its background, B_a, as shown in Eq. (2).

$$C_a = B_a - b_a/B_a. \tag{2}$$

The test object contrast, C_t, is defined in terms of the brightness of the image of one element of the test object, B_t, and the brightness of the image of the background, B_b, as shown in Eq. (3).

$$C_t = B_t/B_b + B_t = TB_a - Tb_a/TB_a = C_a. \tag{3}$$

As is indicated, these brightnesses are related to the apparent brightness of the test object and its background by the transmission factor, T, of the optical system under consideration.

The diffraction pattern contrast, C_d, is defined in terms of the maximum and minimum brightnesses of the superposed diffraction patterns of the test object elements, B_{max} and B_{min}, in the absence of a luminous background. This is shown

TABLE I. The brightness of diffraction patterns for point and line test objects for a circular aperture one millimeter in diameter.

Angle radians $\times 10^5$	Brightness for a point	Brightness for a line	Angle radians $\times 10^5$	Brightness for a point	Brightness for a line
0	100.00	100.00	105.996	.85	1.64
3.533	99.00	97.69	109.529	.52	1.37
7.066	96.06	95.62	113.062	.32	1.09
10.6000	91.34	90.15	116.596	.14	.91
14.133	85.03	85.34	120.129	.04	.86
17.666	77.46	76.75	123.662	.0002	.90
21.199	68.97	67.64	127.195	.02	1.04
24.732	59.94	58.43	130.728	.09	1.09
28.266	50.75	49.10	134.262	.18	1.14
31.799	41.74	39.77	137.795	.27	1.18
35.332	33.26	31.76	141.328	.34	1.24
38.865	25.54	23.86	144.861	.40	1.15
42.398	18.79	17.92	148.394	.42	1.12
45.932	13.12	12.90	151.928	.40	.91
49.465	8.57	8.35	155.461	.36	.73
52.998	5.11	5.76	158.994	.30	.65
56.531	2.67	3.76	162.527	.22	.41
60.064	1.11	3.00	166.060	.15	.26
63.598	.28	2.99	169.594	.08	.24
67.131	.004	3.05	173.127	.04	.26
70.664	.11	3.44	176.662	.008	.28
74.197	.44	3.73	180.193	.0002	.33
77.730	.85	4.01	183.726	.01	.36
81.264	1.25	4.16	187.260	.04	.42
84.797	1.55	4.17	190.793	.07	.45
88.330	1.72	4.04	194.326	.10	.48
91.863	1.74	3.70	197.859	.13	.47
95.396	1.64	3.09	201.392	.15	.45
98.930	1.43	2.56	204.926	.16	.43
102.463	1.15	2.24	208.459	.155	.32

TABLE II. Diffraction pattern data for point sources.

α	B_{max}	B_{min}	Difference ($B_{max}-B_{min}$)	α	B_{max}	B_{min}	Difference ($B_{max}-B_{min}$)
49	1.199	1.198	.001	81	1.011	.448	.563
50	1.174	1.172	.002	82	1.013	.430	.583
51	1.150	1.147	.003	83	1.014	.410	.604
52	1.126	1.122	.004	84	1.015	.392	.623
53	1.108	1.098	.010	85	1.016	.375	.641
54	1.085	1.070	.015	86	1.017	.358	.659
55	1.070	1.048	.022	87	1.018	.338	.680
56	1.056	1.022	.034	88	1.019	.322	.697
57	1.040	1.000	.040	89	1.019	.305	.714
58	1.030	.975	.055	90	1.020	.292	.728
59	1.022	.952	.070	91	1.020	.278	.742
60	1.012	.926	.086	92	1.019	.264	.755
61	1.010	.900	.110	93	1.019	.249	.770
62	1.009	.880	.129	94	1.018	.234	.784
63	1.006	.856	.150	95	1.017	.220	.797
64	1.004	.830	.174	96	1.016	.209	.807
65	1.003	.802	.201	97	1.015	.195	.820
66	1.002	.775	.227	98	1.014	.182	.832
67	1.001	.752	.249	99	1.013	.170	.843
68	1.000	.730	.270	100	1.012	.160	.852
69	1.000	.707	.293	110	1.005	.070	.935
70	1.000	.680	.320	120	1.001	.024	.977
71	1.001	.658	.343	130	1.000	.002	.998
72	1.002	.632	.370	140	1.002	.006	.996
73	1.003	.610	.393	150	1.005	.008	.997
74	1.004	.590	.414	160	1.003	.022	.981
75	1.005	.570	.435	170	1.000	.012	.988
76	1.006	.546	.460	180	1.000	.006	.994
77	1.007	.526	.481	190	1.001	.005	.996
78	1.008	.502	.506	200	1.004	.008	.996
79	1.009	.488	.521	210	1.004	.006	.998
80	1.010	.468	.542				

in Eq. (4).

$$C_d = B_{max} - B_{min}/B_{max}. \quad (4)$$

The image contrast, C_i, is the contrast of the superposed diffraction patterns in the presence of the image of a luminous background at the surface of the photosensitive receptor used with the optical system under consideration. This is shown in Eq. (5).

$$C_i = B_{max} - B_{min}/B_{max} + B_b. \quad (5)$$

By means of Eq. (5), it is possible to select the image contrast for a given angular separation of the elements of the test object by adjusting the brightness of the image of the background. This adjustment automatically fixes the apparent contrast of the test object as indicated by Eq. (2). This provides a means of relating certain quantities of interest. In particular, the coefficient of specific resolution may be related to the apparent contrast for any desired image contrast. Such relations for point and line test objects are shown in Figs. 2 and 3. The numerical values for the brightness of the diffraction patterns were derived from theoretical data of Airy[1] and André.[2] These derived data are shown in Table I for wavelength of 5550A. The brightness of the superposed diffraction patterns in the absence of a luminous background, for different angular separations of the test object elements, are shown in tables 2 and 3, and Figs. 4–7.

APPLICATIONS

There are a number of applications for such data as are presented in Figs. 2 and 3. Three of these are mentioned:

1. The contrast of the image of a given test object at the limit of resolution may be determined for a given optical system. This requires a knowledge of the minimum separable angle of the test object elements and the apparent contrast of the test object with respect to its luminous background. This application is illustrated in Table IV. In this case the contrast of the image on the retina of the human eye is shown at the limit of resolution for Foucault test objects (lines) using a high power telescope.

2. The contrast discrimination required for the photo-sensitive receptor to be used with a par-

TABLE III. Diffraction pattern data for line sources.

α	B_{max}	B_{min}	Difference ($B_{max}-B_{min}$)	C_d%	α	B_{max}	B_{min}	Difference ($B_{max}-B_{min}$)	C_d%
49.0	—	—			75	1.040	.528	.512	49.23
49.5	1.160	1.159	.001		76	1.040	.510	.530	50.96
50.0	1.150	1.142	.008		77	1.040	.490	.550	52.88
50.5	1.150	1.140	.010		78	1.041	.470	.571	54.85
51.0	1.125	1.120	.005	.444	79	1.042	.452	.590	56.62
51.5	1.113	1.105	.008	.712	80	1.043	.430	.613	58.77
52.0	1.103	1.092	.011	.997	81	1.044	.413	.631	60.44
52.5	1.091	1.081	.010	.917	82	1.044	.396	.648	62.07
53.0	1.081	1.065	.016	1.48	83	1.045	.380	.665	63.64
53.5	1.073	1.054	.019	1.77	84	1.044	.364	.680	65.13
54	1.057	1.031	.026	.9524	85	1.044	.350	.694	66.48
55	1.045	1.007	.038	2.97	86	1.044	.333	.711	68.10
56	1.040	.985	.055	5.29	87	1.043	.320	.723	69.32
57	1.045	.958	.087	8.33	88	1.042	.300	.742	71.21
58	1.033	.933	.100	9.68	89	1.041	.288	.753	72.33
59	1.031	.905	.126	12.22	90	1.040	.276	.764	73.46
60	1.030	.880	.150	14.56	95	1.033	.210	.823	79.67
61	1.030	.850	.180	17.48	100	1.025	.160	.865	84.39
62	1.030	.830	.200	19.42	110	1.011	.090	.921	91.09
63	1.030	.800	.230	22.33	120	1.008	.060	.948	94.05
64	1.030	.780	.250	24.27	130	1.011	.060	.951	94.07
65	1.030	.757	.273	26.50	140	1.013	.070	.943	93.09
66	1.030	.732	.298	28.93	150	1.010	.080	.930	92.08
67	1.030	.708	.322	31.26	160	1.008	.087	.921	91.37
68	1.031	.680	.351	34.04	170	1.003	.050	.953	95.01
69	1.030	.660	.371	35.98	180	1.003	.040	.963	96.01
70	1.035	.638	.397	38.36	190	1.004	.040	.964	96.02
71	1.035	.612	.423	40.87	200	1.003	.050	.953	95.01
72	1.037	.593	.444	42.82	210	1.000	.035	.965	96.50
73	1.038	.568	.470	45.28	220	1.000	.025	.975	97.50
74	1.038	.550	.488	47.01	230	1.000	.020	.980	98.00

[1] G. B. Airy, Camb. Phil. Trans. 283 (1834).
[2] C. André, Ann. de l'Ecole Norm. 275 (1876).

ticular optical system performing a given resolution task can be determined.

3. The contrast discrimination of photosensitive receptors used in resolution measurements can be determined for a variety of conditions. This can be done by determining the minimum separable angle for an optical system of known characteristics, using the photosensitive receptor of interest with it. For example, it is evident from Table IV that the contrast discrimination of the retina of the human eye is approximately two percent at an adaptation level of 10 footlamberts for the particular over-all subtense of the test object used.

NOTES ON THEORETICAL RESOLVING POWER

There appears to be considerable confusion concerning what is commonly called "the theoretical resolving power of optical systems." This confusion arises primarily as a result of the arbitrary (and frequently misinterpreted) criterion for resolution established by Airy in 1834. This criterion was (and still is) useful in discussing the diffraction of light by a circular aperture from a mathematical point of view. There are, however, other criteria which may be useful in making physical measurements of resolving power. For example, the limit might be defined as the angular separation of the elements of the test object at which the contrast threshold of the photosensitive receptor is reached for the superposed diffraction patterns of the elements of the test object. This perhaps is the most used criterion. A second criterion might be taken to be the angle subtended by the elements of the test object when the number of the elements of the test object can no longer be specified. In the former criterion using the human eye and point test object elements of equal brightness, the limit would occur at about 55.6×10^{-5} milli-

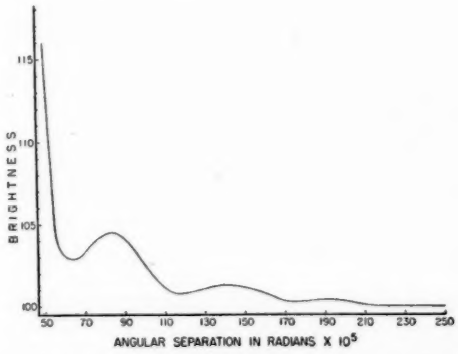


FIG. 4. The maximum brightness of superposed diffraction patterns for a circular aperture of one mm *vs.* angular separation of luminous points of equal brightness.

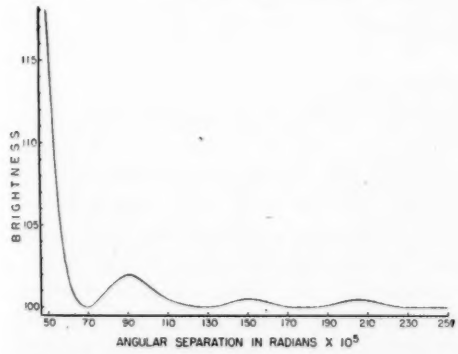


FIG. 5. The maximum brightness of superposed diffraction patterns for a circular aperture of one mm *vs.* angular separation of luminous lines of equal brightness.

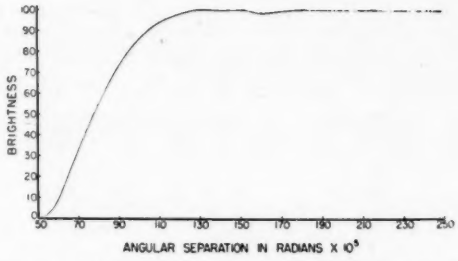


FIG. 6. Difference between the maximum and the minimum brightness of the superposed diffraction patterns for a circular aperture of one mm *vs.* angular separation of luminous points of equal brightness.

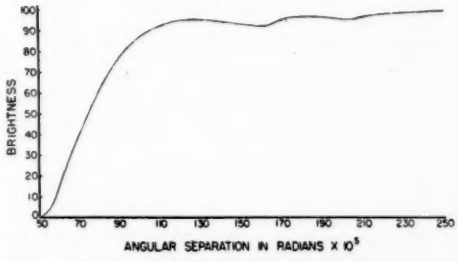


FIG. 7. Difference between the maximum and the minimum brightness of the superposed diffraction patterns for a circular aperture of one mm *vs.* angular separation of luminous lines of equal brightness.

TABLE IV. The brightness contrast of the image on the retina for Foucault test objects of different apparent contrast at the limit of resolution of an 80× telescope and a brightness level of 10 footlamberts.

Apparent contrast (percent)	Coefficient of specific resolution (mm −rad ×10⁵)	Image contrast at limit of resolution (percent)
100	54.6	3.0
63	55.6	3.0
44	55.8	2.5
31	56.2	2.2
23	56.7	2.0
21	57.4	2.0
12	58.1	1.5
9	60.0	1.5
6	64.3	1.5
4	76.5	2.1
3	99.0	2.5

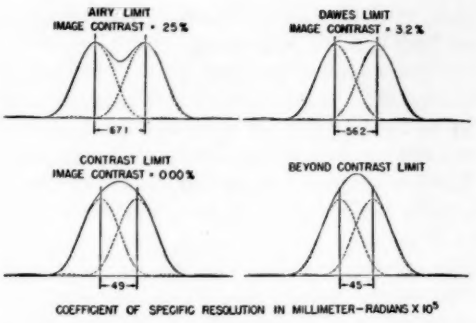


FIG. 8. The coefficient of specific resolution for various criteria of resolution using point test objects and a circular aperture.

meter-radian (according to Dawes[3]) and in the latter case may be as high as 67.1×10^{-5} millimeter-radian (using Airy's criterion). These criteria are illustrated in Fig. 8. Still other criteria for theoretical resolution might be used. For example, the criterion could depend only upon whether or not a decision could be made as to the possible presence of more than one element of the test object in the diffraction pattern. In this case the limiting angle may be very close to zero.

[3] W. R. Dawes, Me. Roy. Astr. Soc. **35**, 154 (1867).

ACKNOWLEDGMENT

The authors are indebted to Mr. Frank Tidona and Mr. Andrew Vrabel for assistance with the calculations and preparation of the figures presented in this paper.

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 37, NUMBER 7 JULY, 1947

The Reflectivity and Transmissivity of Multiple Thin Coatings

W. WEINSTEIN
Research Department, Adam Hilger Ltd., London, England

(Received February 13, 1947)

General expressions for the transmissivity and reflectivity of multi-layer coatings of any number of components, for light polarized in any way and incident at any angle, are derived.

INTRODUCTION

WITHIN the last ten years, the use of thin coatings of transparent materials on optical elements has become widespread; "thin" here implies a thickness of the order of magnitude of a wave-length of light. The purpose of such coatings is to modify the reflection characteristics of the surface without loss of energy by absorption; the most usual case is that of "blooming," where one film (occasionally two) is used to reduce surface reflectivity to a minimum, but several films, alternating with high and low refractive indices, have been used to increase the reflectivity, applications being to beam-splitting devices where metallic films would cause an undesirable loss of energy, and to filters for special purposes. The theory of such multiple coats has been treated for the cases of one, two and three films with normal incidence. In this note, the general case of any number of films, with plane waves incident at any angle and polarized in any azimuth, is treated. The effect of absorption is also considered.

The starting point is the set of equations ex-

pressing the boundary conditions which must hold between the various electromagnetic radiations at the film interfaces, as in a paper by Mooney.[1] In the present note, these are written down for any number of films, with plane waves incident at any angle, as stated above, and two methods of solution are developed, each giving the ratios between the incident and reflected amplitudes and the incident and transmitted amplitudes. By one method, explicit formulae are obtained for the transmittance and reflectance, and by the other the solutions are expressed in matrix notation; the latter form, while not convenient for exhibiting the general properties of the multiple coatings, is better adapted for numerical calculation.

EQUATIONS FOR THE BOUNDARY CONDITIONS

Case I. The Electric Intensity Vector Parallel to the Plane of Incidence

It will be assumed that there are $p+1$ media, and quantities belonging to these will have subscripts 0 to p. A right-handed coordinate system will be used, with the x axis normal to the film surface; plane waves incident in the xy-plane on the coating from medium p are considered;[2] the origin is in the interface between medium p and medium $p-1$, and other interfaces have positive x coordinates (Fig. 1).

Let:

n_k = refractive index of medium k ($k=0, 1, \cdots p$).
h_k = thickness of medium k ($k=1, 2, \cdots p-1$).
E_p^+ = amplitude of electric intensity vector in incident beam.
E_p^- = amplitude of electric intensity vector in reflected beam.
E_0^+ = amplitude of electric intensity vector in transmitted beam.

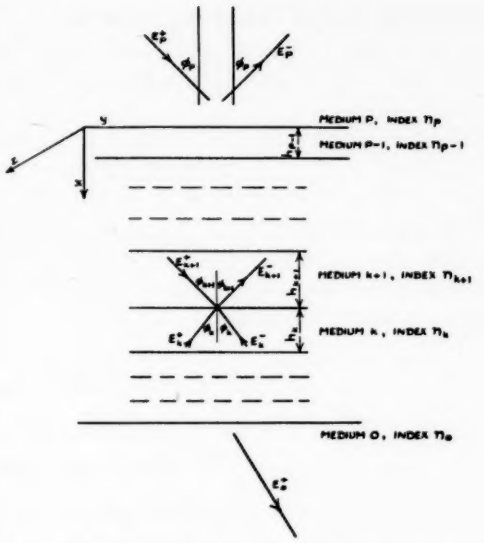


FIG. 1. The figure illustrates the coordinate system and notation used. The z-axis projects upwards out of the plane of the paper. The arrows attached to E_k^+, E_k^-, etc., are on the normal to the wave fronts of these radiations, and indicate the direction of travel of the wave fronts. (To read the figure correctly, E_k^+ and E_k^- should be interchanged and the arrow on E_k^+ reversed.)

E_k^+ = amplitude of electric intensity vector in beam traveling in positive direction in medium k ($k=1, 2, \cdots p-1$).
E_k^- = amplitude of electric intensity vector in beam traveling in negative direction in medium k ($k=1, 2, \cdots p-1$).
ϕ_k = angle of incidence in the positive direction in medium k ($k=0, 1, \cdots p$).
Y_k^+ = y component of E_k^+ ($k=0, 1, \cdots p$).
Y_k^- = y component of E_k^- ($k=0, 1, \cdots p$).
γ_k^+ = z component of magnetic intensity vector corresponding to E_k^+.
γ_k^- = z component of magnetic intensity vector corresponding to E_k^-.

For plane waves, polarized with the electric vector in the $x-y$ plane:

$$\left.\begin{aligned} Y_k^+ &= \cos\phi_k E_k^+ \exp\frac{2\pi i}{\lambda}\{n_k(x\cos\phi_k + y\sin\phi_k) - ct\} \\ Y_k^- &= \cos\phi_k E_k^- \exp\frac{2\pi i}{\lambda}\{n_k(-x\cos\phi_k + y\sin\phi_k) - ct\} \end{aligned}\right\} \cdots, \tag{1}$$

[1] R. L. Mooney, J. Opt. Soc. Am. **25**, 574 (1945).
[2] Remark on notation: the light is taken as incident from medium p and not, as might be expected, from medium 0, because, as will be seen later, the solution is obtained by successive substitutions beginning with terms containing the subscript appropriate to the medium into which the transmitted wave emerges; thus another film can be added by making one more substitution at the end of the calculation, which is done by simply changing p into $p+1$ in the general formula.

and using Maxwell's equations we have

$$\left.\begin{aligned}\gamma_k^+ &= n_k E_k^+ \exp\frac{2\pi i}{\lambda}\{n_k(x\cos\phi_k + y\sin\phi_k) - ct\}\\ \gamma_k^- &= -n_k E_k^- \exp\frac{2\pi i}{\lambda}\{n_k(-x\cos\phi_k + y\sin\phi_k) - ct\}\end{aligned}\right\}. \tag{2}$$

(The amplitudes $E_k^{\pm}$ can be complex quantities, so that the phase constants are included in them.)

The relevant boundary conditions are that the tangential components of the electric and magnetic intensity vectors shall be continuous at each interface; (the boundary conditions for the normal components merely express Snell's law).

Let:

$$\begin{aligned}&v_k = (2\pi/\lambda)n_k\cos\phi_k\cdots(k=0,\,1,\,\cdots p),\\ &x_k^+ = \cos\phi_k E_k^+\exp\{iv_k(h_{p-1}+h_{p-2}+\cdots+h_k)\}\cdots(k=1,\,2,\,\cdots p-1),\\ &x_k^- = \cos\phi_k E_k^-\exp\{-iv_k(h_{p-1}+h_{p-2}+\cdots+h_k)\}\cdots(k=1,\,2,\,\cdots p-1),\\ &x_0^+ = \cos\phi_0 E_0^+\exp\{iv_0(h_{p-1}+h_{p-2}+\cdots+h_1)\},\\ &x_p^+ = \cos\phi_p E_p^+,\\ &x_p^- = \cos\phi_p E_p^-,\\ &u_{pk} = n_k\sec\phi_k\cdots(k=0,\,1,\,\cdots p).\end{aligned}$$

If the two boundary conditions for each interface are written down, factors of the form $\exp(2\pi i/\lambda)(n_k y\sin\phi_k - ct)$ cancelled, and the above substitutions made, the following set of $2p$ simultaneous linear equations is obtained, involving homogeneously the $2p+1$ unknowns $x_0^+, x_k^{\pm}, (k=1,2,\cdots p)$.

$$\left.\begin{aligned}x_1^+ + x_1^- - x_0^+ &= 0\\ u_1x_1^+ - u_1x_1^- - u_0x_0^+ &= 0\\ \left.\begin{aligned}x_{k+1}^+ + x_{k+1}^- - e^{-iv_kh_k}x_k^+ - e^{iv_kh_k}x_k^- &= 0\\ u_{k+1}x_{k+1}^+ - u_{k+1}x_{k+1}^- - u_ke^{-iv_kh_k}x_k^+ + u_ke^{iv_kh_k}x_k^- &= 0\end{aligned}\right\}(k=1,\,2,\,\cdots p-1)\end{aligned}\right\}. \tag{3}$$

These must be solved for x_p^+ and x_p^- in terms of x_0^+, and the transmissivity T_p and reflectivity R_p of the coating are then given by

$$T_p = \frac{n_0\cos^2\phi_p}{n_p\cos^2\phi_0}\left|\frac{x_0^+}{x_p^+}\right|^2;\quad R_p = \left|\frac{x_p^-}{x_p^+}\right|^2, \tag{4}$$

where $|A|$ denotes the modulus of the complex quantity A.

Case II. The Electric Intensity Vector Perpendicular to the Plane of Incidence

The procedure here is similar to that in Case I. Let $Z_k^{\pm}$ be the z-components of the electric intensity vectors in the positive and negative directions in medium k (there are no x or y-components). Let $\beta_k^{\pm}$ be the y-components of the corresponding magnetic intensity vectors. As before,

$$\left.\begin{aligned}Z_k^+ &= E_k^+\exp(2\pi i/\lambda)\{n_k(x\cos\phi_k + y\sin\phi_k) - ct\}\\ Z_k^- &= E_k^-\exp(2\pi i/\lambda)\{n_k(-x\cos\phi_k + y\sin\phi_k) - ct\}\end{aligned}\right\},$$

and

$$\left.\begin{aligned}\beta_k^+ &= -n_k\cos\phi_k E_k^+ \exp\frac{2\pi i}{\lambda}\{n_k(x\cos\phi_k+y\sin\phi_k)-ct\}\\ \beta_k^- &= n_k\cos\phi_k E_k^- \exp\frac{2\pi i}{\lambda}\{n_k(-x\cos\phi_k+y\sin\phi_k)-ct\}\end{aligned}\right\}.$$

Let:

$$\begin{aligned}y_k^+ &= E_k^+\exp\{iv_k(h_{p-1}+h_{p-2}+\cdots+h_k)\}\quad (k=1,2,\cdots p-1),\\ y_k^- &= E_k^-\exp\{-iv_k(h_{p-1}+h_{p-2}+\cdots+h_k)\}\quad (k=1,2,\cdots p-1),\\ y_0^+ &= E_0^+\exp\{iv_0(h_{p-1}+h_{p-2}+\cdots+h_1)\},\\ \omega_k &= n_k\cos\phi_k\quad (k=0,1,\cdots p).\end{aligned}$$

The boundary conditions are the same as before, and on making the above substitutions we obtain a set of equations for the y's,

$$\left.\begin{aligned}y_1^+ + y_1^- - y_0^+ &= 0\\ \omega_1 y_1^+ - \omega_1 y_1^- - \omega_0 y_0^+ &= 0\\ y_{k+1}^+ + y_{k+1}^- - e^{-iv_k h_k} y_k^+ - e^{iv_k h_k} y_k^- &= 0\\ \omega_{k+1} y_{k+1}^+ - \omega_{k+1} y_{k+1}^- - \omega_k e^{-iv_k h_k} y_k^+ + \omega_k e^{iv_k h_k} y_k^- &= 0\end{aligned}\right\}(k=1,2,\cdots p-1)\Bigg\}.$$

It can be seen that these equations are simply the result of putting $y_k^{\pm}$ for $x_k^{\pm}$ and ω_k for u_k in Eqs. (3), and hence the same method of solution will serve in both cases. The transmissivity T_p' and reflectivity R_p' are then given by

$$T_p' = \frac{n_0}{n_p}\left|\frac{y_0^+}{y_p^+}\right|^2,\quad R_p' = \left|\frac{y_p^-}{y_p^+}\right|^2. \tag{5}$$

SOLUTION OF EQUATIONS (3)

Each pair of equations in (3) can be solved to give expressions for x_{k+1}^+ and x_{k+1}^- in terms of x_k^+ and x_k^-, giving

$$\left.\begin{aligned}x_{k+1}^+ &= \frac{1}{2u_{k+1}}\{(u_{k+1}+u_k)e^{-iv_k h_k}x_k^+ + (u_{k+1}-u_k)e^{+iv_k h_k}x_k^-\}\cdots(k=1,2,\cdots p-1)\\ x_{k+1}^- &= \frac{1}{2u_{k+1}}\{(u_{k+1}-u_k)e^{-iv_k h_k}x_k^+ + (u_{k+1}+u_k)e^{iv_k h_k}x_k^-\}\cdots(k=1,2,\cdots p-1)\\ x_1^+ &= \frac{1}{2u_1}(u_1+u_0)x_0^+\\ x_1^- &= \frac{1}{2u_1}(u_1-u_0)x_0^+\end{aligned}\right\}. \tag{6}$$

These constitute a set of recurrence formulas for the x_k^+'s and x_k^-'s, and they can be used in succession to give x_p^+ and x_p^- in terms of x_0^+, by substituting for x_1^+ and x_1^- in the formulas for x_2^+ and x_2^-, then substituting for x_2^+ and x_2^- in the formulas for x_3^+ and x_3^-, and so on.

If this is done it is easily seen that the general formulas for x_p^+ and x_p^- are given as follows:

$$\begin{aligned}x_p^+ &= (x_0^+/2^p u_1 u_2\cdots u_p)\{H_1+H_2+\cdots+H_{2p-1}\},\\ x_p^- &= (x_0^+/2^p u_1 u_2\cdots u_p)\{K_1+K_2+\cdots+K_{2p-1}\},\end{aligned}$$

where each of the H's and K's is of the form

$$(u_p \pm u_{p-1})(u_{p-1} \pm u_{p-2}) \cdots (u_1 \pm u_0) \exp\{\pm iv_{p-1}h_{p-1} \pm iv_{p-2}h_{p-2} \pm \cdots \pm iv_1h_1\}$$

with the ambiguous signs suitably determined. The sign of iv_kh_k in both H_j and K_j is the sign of -1 raised to the power $1+[j^{-1}/2^{k-1}]$, where $[j^{-1}/2^{k-1}]$ denotes the largest integer contained in $(j^{-1}/2^{k-1})$. The sign of u_k in the factor $(u_{k+1} \pm u_k)$ in H_j is the sign of -1 raised to the power $[(j+2^{k-1}-1)/2^k]$, where the square brackets again denote the largest integer contained in the quantity inside. The signs in the coefficients $(u_{k+1} \pm u_k)$ in K_j are the same as those in the corresponding term, H_j, except for the first factors in each term, namely $(u_p \pm u_{p-1})$; in these the signs are the opposite of the corresponding signs in H_j.

(The meaning of these rules is that the signs of iv_kh_k in the sequence $H_1, H_2, \cdots H_{2p-1}$ consist of alternate successions of 2^{k-1} minus signs and 2^{k-1} plus signs, beginning with the minus signs, and that the signs in the factors $(u_{k+1} \pm u_k)$ commence with 2^{k-1} plus signs, after which come alternate successions of 2^k minus signs and 2^k plus signs; similarly for $K_1, K_2, \cdots K_{2p-1}$.)

It is then seen that $2^{2p}u_1^2u_2^2 \cdots u_p^2 |x_p^+/x_0^+|^2$, which is equal to

$$2^{2p}u_1^2u_2^2 \cdots u_p^2 \left(\frac{n_p \cos^2\phi_0}{n_0 \cos^2\phi_p} T_p\right)^{-1},$$

is of the form

$$\sum_{j=1}^{2p-1} A_j^2 + 2 \sum_{j \neq k} A_jA_k \cos(b_j - b_k),$$

when the A's are the products $(u_p \pm u_{p-1}) \cdots (u_1 \pm u_0)$ and the b's are the sums $\pm v_{p-1}h_{p-1} \pm v_{p-2}h_{p-2} \pm \cdots \pm v_1h_1$; i.e.,

$$T_p = \frac{n_0 \cos^2\phi_p}{n_p \cos^2\phi_0} \left(\frac{2^{2p}u_1u_2 \cdots u_p}{\sum_{j=1}^{2p-1} A_j^2 + 2 \sum_{j \neq k} A_jA_k \cos(b_j - b_k)} \right), \tag{7}$$

with a similar expression, involving ω's instead of u's for T_p'. R_p and R_p' can be obtained similarly, but it is simpler to use Eq. (14), below.

An alternative method of solution, more convenient for computation, is as follows:

The Eqs. (6) can be put in either of the matrix forms:

$$\begin{bmatrix} x_{k+1}^+ \\ x_{k+1}^- \end{bmatrix} = \frac{1}{2u_{k+1}} \begin{bmatrix} (u_{k+1}+u_k)e^{-iv_kh_k} & (u_{k+1}-u_k)e^{iv_kh_k} \\ (u_{k+1}-u_k)e^{-iv_kh_k} & (u_{k+1}+u_k)e^{iv_kh_k} \end{bmatrix} \cdot \begin{bmatrix} x_k^+ \\ x_k^- \end{bmatrix}, \tag{8}$$

$$[x_{k+1}^+ \quad x_{k+1}^-] = \frac{1}{2u_{k+1}} [x_k^+ \quad x_k^-] \cdot \begin{bmatrix} (u_{k+1}+u_k)e^{-iv_kh_k} & (u_{k+1}-u_k)e^{-iv_kh_k} \\ (u_{k+1}-u_k)e^{iv_kh_k} & (u_{k+1}+u_k)e^{iv_kh_k} \end{bmatrix}. \tag{9}$$

Denoting the square matrix on the right-hand side of (8) by M_k, these can be written as

$$\begin{bmatrix} x_{k+1}^+ \\ x_{k+1}^- \end{bmatrix} = \frac{1}{2u_{k+1}} M_k \begin{bmatrix} x_k^+ \\ x_k^- \end{bmatrix}, \tag{8'}$$

$$[x_{k+1}^+ \quad x_{k+1}^-] = \frac{1}{2u_{k+1}} [x_k^+ \quad x_k^-] M_k', \tag{9'}$$

where M_k' is the transpose of M_k.

Hence

$$\begin{bmatrix} x_p^+ \\ x_p^- \end{bmatrix} = \frac{1}{2^p u_1 u_2 \cdots u_p} M_{p-1} M_{p-2} \cdots M_1 \begin{bmatrix} (u_1+u_0)x_0^+ \\ (u_1-u_0)x_0^+ \end{bmatrix}, \tag{10}$$

$$[x_p^+ \quad x_p^-] = \frac{1}{2^p u_1 u_2 \cdots u_p} [(u_1+u_0)x_0^+ \quad (u_1-u_0)x_0^+] M_1' M_2' \cdots M_{p-1}'. \tag{11}$$

Pre-multiplying the complex conjugate of (11) by (10)

$$\begin{bmatrix} \left|\frac{x_p^+}{x_0^+}\right|^2 & \frac{x_p^+ \bar{x}_p^-}{|x_0^+|^2} \\ \frac{\bar{x}_p^+ x_p^-}{|x_0^+|^2} & \left|\frac{x_p^-}{x_0^+}\right|^2 \end{bmatrix} = \frac{M_{p-1} M_{p-2} \cdots M_1}{2^{2p} u_1^2 u_2^2 \cdots u_p^2} \begin{bmatrix} (u_1+u_0)^2 & (u_1^2-u_0^2) \\ (u_1^2-u_0^2) & (u_1-u_0)^2 \end{bmatrix} \bar{M}_1' \bar{M}_2' \cdots \bar{M}_{p-1}'; \tag{12}$$

the leading diagonal elements can then be substituted in (4) to give T_p and R_p. As stated above, the same method is available for finding T_p' and R_p'.

If the incident beam contains intensity fractions C_1 and C_2 polarized with the electric vector in and normal to the plane of incidence respectively, (for unpolarized light $C_1 = C_2$), the total transmissivity T and reflectivity R are given by

$$T = (C_1 T_p + C_2 T_p')/(C_1+C_2); \quad R = (C_1 R_p + C_2 R_p')/(C_1+C_2).$$

ABSORBING MEDIA

If there is any appreciable absorption this may be taken account of in the usual way by substituting for the refractive indices n_k the complex refractive indices $n_k + i\alpha_k$; here α_k is the absorption index, such that after traversing a distance of one wave-length, the amplitude is reduced in the ratio $\exp(-2\pi\alpha_k)$. α is related to the extinction coefficient D by the formula

$$4\pi\alpha/\lambda = D \log_e 10. \tag{13}$$

The solution can be carried out as before, and it will then be found that real exponential factors appear in the terms of the expressions for T and R.

DISCUSSION OF THE RESULTS

It is obvious that the expressions for the reflected and transmitted amplitudes, and still more those for R and T, become very complicated when more than two or three layers are considered, and if there is no absorption it is simplest then to calculate T directly and obtain R from the relation

$$R + T = 1. \tag{14}$$

If it is desired to obtain expressions for T and R explicitly as functions of λ and the various thicknesses, then the first method of solution must be used. The reciprocal of T will contain $2^{p-2}(2^{p-1}-1)+1$ terms in the most general case, but usually it will happen that not all the layers have different refractive indices; in fact in making, for example, beam splitters with $T = \frac{1}{2}$, it would be assumed that the layers are symmetrical with respect to the refractive indices and thicknesses about a central plane; then a number of terms will be found to vanish and the complexity of the expressions will be considerably reduced.

If on the other hand, the reflectivity (or transmissivity) is required for a given set of film thicknesses and refractive indices, it will be quicker to use the matrix expression (12); here the multiplications should be carried out beginning at the center, pre-multiplying the central matrix by M_1, then post-multiplying by $\bar{M}_1'$ and so on; in this way a real product is obtained at every second step, thus simplifying the numerical work.

ACKNOWLEDGMENT

Thanks are due the directors of Messrs. Adam Hilger Ltd. for permission to publish this note.

Compensation of Flexure in Range Finders and Sighting Instruments*

G. W. MOFFITT
Ridgewood, New Jersey

(Received March 21, 1947)

A method has been developed for the automatic compensation of the errors caused by bending in self-contained base range finders, especially the error arising from one-sided heating —as from sunshine, or from gravity when the elevation of the line of sight is changed. The conditions may be met in instruments of existing types and sizes. At the same time the construction is simplified and the reliability improved. The "optical tube" may be eliminated entirely in some types of range finder. These results are obtained through the use of objectives composed of two widely separated components so related to each other in power and position that the usual lateral displacements of the image on the focal plane are compensated by corresponding lateral displacements of the second principal point of the compensating objectives. The simple case in which the axis of a range finder bends in the form of a circular arc is discussed. Compensation of this kind may also be used in the designing of sighting telescopes.

IN bringing the self-contained base range finder to its present state of development much effort has been expended on the problems arising from the presence of flexure in the structure of the instrument. Heretofore, nearly all designers have tried to make the supporting structure, especially the optical tube, as rigid as possible in the hope that flexure could be kept small enough to be harmless. Designs have sometimes been greatly complicated in this way, but flexure still remains in sufficient amount to be troublesome. In comparatively recent years some attempts have been made to simplify the structure and improve the results by using the principles of compensation. For example, it has been shown[1] that the use of tele-objectives in place of ordinary telescope objectives reduces the effects of flexure on the readings of range. However, if one analyzes the problem it will be found that only slight improvement is possible in this way.

A few years ago the writer returned for a while to a consideration of problems in range finder design and took up the question of flexure in a more general way to see if any arrangement of lenses can exist that will result in practically complete compensation and, if so, to determine the conditions that must be fulfilled. It was found that control of compensation is possible through the use of objective systems composed of an anterior negative lens, or lens group, and a posterior positive lens, or lens group, the two lenses being so proportioned and spaced that the desired result is obtained. Although flexure from one-sided heating was in mind, the method is also applicable to the compensation of bending from other causes.

It is well known that heating one side of a tubular structure causes the ends of the tube to bend away from the source of the heat. Although shielded to a greater or less extent by the outer parts of the instrument, the optical tube of an ordinary range finder will experience this bending to some extent, with consequent errors in the determination of range. Figure 1 shows schematically the geometry of the optical tube in an ordinary range finder where f is the equivalent

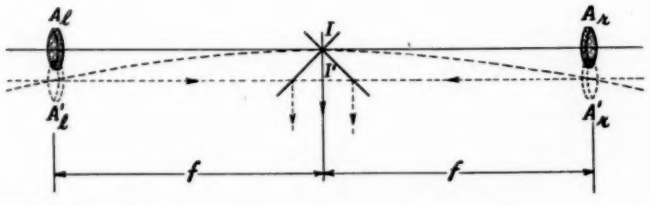


FIG. 1. Illustrating the effect of flexure in the optical tube of a conventional range finder with ordinary telescope objectives.

* Presented at the meeting of the Optical Society of America in New York, New York, October, 1946.
[1] Artur Pulz, U. S. Patent No. 2,008,516, July 16, 1935.

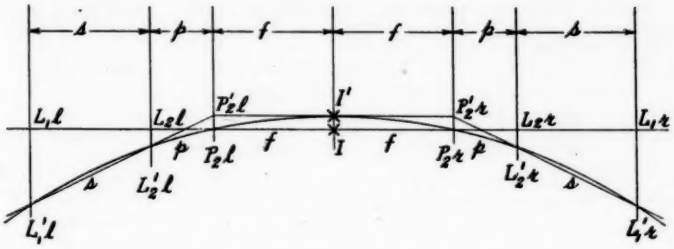


FIG. 2. Illustrating the compensation of flexure when an anterior negative lens, followed at a distance by a posterior positive lens, is substituted for each telescope objective of a conventional range finder.

focal length of the objectives A_l and A_r with common focal plane at the middle—here represented by a cross. Whenever such an optical tube becomes curved, through asymmetrical heating or any other cause, the second principal points of the objectives will be thrown out of alignment with the coincidence prism, and readings of range will then be in error. The dotted portion of Fig. 1 shows the situation greatly exaggerated for purposes of illustration. In effect, the objectives have moved over to A_l' and A_r'. The portions of the image of the target seen above and below the dividing line are no longer in coincidence but are in error by $2II'/f$ in radian measure.

Figure 2 shows schematically the geometry of an optical tube using objective systems having widely spaced components. L_{1l} and L_{2l}, and L_{1r} and L_{2r}, mark the locations of the negative and of the positive components of the left and of the right objective systems respectively, while P_{2l} and P_{2r} mark the locations of the second principal points. The optical separation of the two components of each objective system is represented by s, while the distance from the second component to the second principal point is designated by p. The equivalent focal length of each objective system is designated by f. In a state of good adjustment and of temperature equilibrium the centers of the various components and of the coincidence prism I are on the straight line $L_{1l}IL_{1r}$. But if a temperature gradient of the kind under consideration develops in the optical tube, the whole system will assume a curved configuration (greatly exaggerated in Fig. 2), the various components moving over to the primed positions. It will be noticed that the second principal points and the coincidence prism have moved over in the same direction with respect to the straight line configuration. Inspection of the figure reveals that it is possible to so proportion s, p, and f with respect to each other that the second principal points and the coincidence prism will remain on a straight line in spite of the bending of the optical tube.

Figure 3 shows the left-hand end of the optical tube with the bending greatly exaggerated. For simplicity let us assume that the axis of the optical tube becomes curved in a circular arc. Of course, the actual form of the curve in any case will depend upon the design and construction of the instrument. But, the actual curve having been determined for a given design, the problem can be worked out to suit that curve in much the same manner as for the circular curve. From the figure we have

$$f^2+r^2=c^2 \tag{1}$$

and

$$c^2=[p+(s/2)]^2+d^2. \tag{2}$$

Combining (1) and (2) gives

$$f^2+r^2=[p+(s/2)]^2+d^2. \tag{3}$$

But

$$r^2=(s^2/4)+d^2. \tag{4}$$

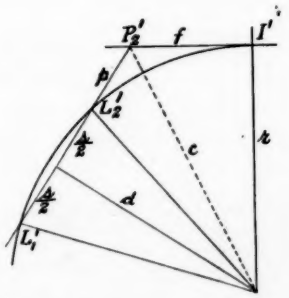


FIG. 3. Illustrating the geometry involved in the compensation of uniform flexure.

TABLE I. Skeletonized tabulation of relative focal lengths and spacings for fully compensated systems in the illustrative case used in the text ($f=+1$).

f_1	f_2	s^2	s	p	B.F.	B.F.$+s$
−0.25	0.3789	0.0500	0.2236	0.8944	1.8944	2.1180
−0.5	0.6055	1.6667	0.4083	0.8166	1.8166	2.2249
−1.0	0.8535	0.5000	0.7071	0.7071	1.7071	2.4142
−2.0	1.0516	1.3333	1.1547	0.5774	1.5774	2.7321
−3.0	1.1250	2.2500	1.5000	0.5000	1.5000	3.0000
−4.0	1.1578	3.2000	1.7889	0.4472	1.4472	3.2361
−8.0	1.1850	···	2.6667	0.3333	1.3333	4.0000

Eliminating r by combining (3) and (4) gives

$$f^2=p(p+s) \tag{5}$$

as the condition sought.

To determine the equivalent focal lengths required for L_1 and L_2 the equations for equivalent focal length and location of principal points are used, as they apply to a system composed of two separated lenses. Let f_1 be the focal length of L_1 and f_2 that of L_2, while s is the optical separation of the two. Further, let f be the equivalent focal length of the combination. Then

$$f=f_1f_2/(f_1+f_2-s) \tag{6}$$

and

$$P_2=-(fs/f_1), \tag{7}$$

positive when P_2 lies to the right of L_2 in Fig. 3. Introducing these values in Eq. (5) gives

$$f^2=-\frac{fs}{f_1}[s-(fs/f_1)] \tag{8}$$

or

$$f=-\frac{s}{f_1}[s-(fs/f_1)] \tag{9}$$

or

$$f=(fs^2/f_1^2)-(s^2/f_1) \tag{10}$$

or

$$s=\pm f_1\left(\frac{f}{f-f_1}\right)^{\frac{1}{2}}. \tag{11}$$

And from Eq. (6) comes

$$f_2=f\left(\frac{s-f_1}{f-f_1}\right). \tag{12}$$

These are the equations with which one may determine completely the focal lengths and the location of the various lenses in order that compensation of bending in a circular arc shall be complete.

Among the possible combinations satisfying these equations that one is of interest here in which a negative lens, or lens group, is followed by a positive lens, or lens group. The conditions are such that the equivalent focal length is positive, an inverted image being formed beyond L_2 and P_2. For unit positive equivalent focal length Eq. (11) becomes

$$s^2=\frac{f_1^2}{1-f_1}, \quad \text{or} \quad s=\pm\frac{f_1}{(1-f_1)^{\frac{1}{2}}}. \tag{13}$$

The data for a number of systems all having an equivalent focal length of $+1$ and satisfying the equations for compensation of bending are given in Table I, all other quantities being expressed in terms of the equivalent focal length.

The upper lines in Table I give lenses having focal lengths too short for the good correction of the aberrations at the required apertures. On the other hand, the lower lines in the table give systems for which the distance from the anterior component to the focal plane may be greater than will fit into instruments of usual dimensions. But between these two extremes there are many combinations quite suitable for instruments of ordinary size and specifications. One such combination is shown to scale in Fig. 4, which represents a stereoscopic instrument of medium size.

Thus from the table one may determine the values that will apply in any given case in which the common focal plane of the objective systems is at, or near, the middle of the instrument, and

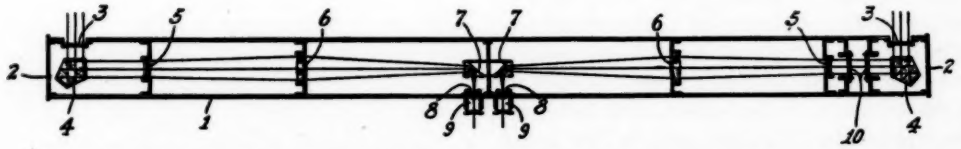


FIG. 4. Central longitudinal section of a stereoscopic range finder fully compensated for uniform flexure.

the bending is uniformly distributed. Usually however, the distance from the second principal point to the middle of the instrument will be somewhat less than the equivalent focal length because a portion of the light path will be folded in various ways within the coincidence prism, or lie on the eyepiece side of it. This can be taken into account by introducing suitable constants in the equations.

Although the results obtained above are for an instrument that bends in the form of a circular arc, it is obvious that the same general method applies to any kind of bending. Once the bent shape is known, the proper equations can be set up and suitable proportions for the system determined.

Not only does the compensation of flexure make possible a greater degree of constancy in the calibration of an instrument but it sometimes opens the way to a marked simplification in the mechanical structure. Sometimes the optical tube may even be dispensed with, as in Fig. 4.

In conclusion, it may be added that the useful field of application for compensation of flexure is not limited to instruments of the self-contained base range finder type. It should also find a sphere of usefulness in the designing of sighting instruments, where it is important to maintain a known relationship between the line of sight and the supports of the instrument, in spite of changes in the angle of elevation or in thermal conditions.

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 37, NUMBER 7 JULY, 1947

Spectrochemical Determination of the Major Constituents of Minerals and Rocks

ASLAK KVALHEIM
*Government Raw Materials Laboratory, Oslo, Norway**

(Received January 29, 1947)

A method is described for the spectrochemical determination of Si, Na, K, Al, Ca, Mg, Fe, and Mn in minerals, rocks, slags, and related substances; the determination of Ti is also discussed. The d.c. carbon arc is employed. One part of powdered sample is mixed with 1 part of $SrCO_3$, which serves as internal standard for the determination of all the elements mentioned except Si, or with 2 parts of quartz and 3 parts of $SrCO_3$ if a more diluted sample is wanted. For the determination of Si the mixture is 1 part of sample +1 part of $BeCO_3$ (internal standard) +3 parts of NaCl. Carbon powder is added to all the mixtures. Three arcings are necessary: for K and Na in the visible and upper ultraviolet region with the sample in the cathode; in the ultraviolet for Si, also with cathode evaporation; another in the ultraviolet for Al, Ca, Mg, Fe, and Mn (and Na) with the sample as anode. The use of different sized electrodes is discussed. Calibration curves are shown, based on the spectra of actual minerals, rocks, slags, etc. having reliable chemical analyses. The relative error is in most cases less than 10 percent.

THE method presented in this paper does not involve any new spectrographic technique. It utilizes the well known direct current arc used in numerous laboratories for quantitative analysis, but it introduces a few comparison elements, which have been found to be especially applicable to the large variety of base substance usually met in a spectrographic rock analysis, and indicates a comparatively simple way of carrying out the determinations.

The method was intended to be used particularly for the determination of major constituents of mineral samples of which only very small amounts might happen to be available (which often occurs in geological problems), and also for rapid analyses of rocks, slags, and similar compounds;[1] it was developed with the following points in mind:

The method should be simple, the samples to be analyzed should be diluted as little as possible with other compounds before the analysis, and chemical treatment avoided; internal standards

* Temporarily associated with Spectrographic Services and Supplies, Inc., Saratoga Springs, New York.

[1] Other methods for the spectrochemical determination of the main components of rocks are published by: H. I. Oshry, J. W. Ballard, and H. H. Schrenk, J. Opt. Soc. Am. **32**, 672–80 (1942); and by J. Eeckhout, Nature **156**, 175 (1945).

should be used, but these should be very few, if possible, only one. The spectral regions should be chosen so that a minimum of exposures would be necessary, and the exposure time should not be long. The method should be applicable to rocks of largely varying composition, thus covering a wide range of concentration, but the analysis should be carried out with a relative error of less than 10 percent and preferably, less than 5 percent.

THE STANDARDS

Since the making of synthetic rock standards is very complicated and difficult, and since mixtures made of the oxides of the elements involved have little significance in this case insofar as the resulting calibration curves obtained from such mixtures must afterward be checked by arcing numerous actual rocks of varied and known composition, the simplest and safest way is to use actual minerals, rocks, slags, etc., with reliable chemical analysis.

The minerals, rocks, slags etc., used as standards for this method are listed in Table I, showing the widely different bases and combinations of elements to which the method is applicable.

Investigations made on a number of rock spectra showed clearly that an analysis could not be made without at least *some* dilution of the rock powder. These investigations also showed that the variations in intensity of the Si lines originating from rocks of approximately the same SiO_2 content were extremely great, thus pointing towards the probable necessity of selecting a special comparison element for Si.

INTERNAL STANDARDS

Many elements in various combinations were tried as intensity standards, but two elements were chosen as decidedly the best, *viz.* Sr and Be.

Sr proved to be very good as comparison element for K, Na, Al, Ca, Mg, Fe, and Mn. Ti might also be included here but is omitted for reasons which are discussed later.

In addition, Sr is present only in very small amounts in the usual rocks, so that disturbances from Sr previously present in the sample will seldom occur. As a comparison element for Si, however, Sr is not suitable, the resulting data being quite confusing.

TABLE I.*

S.H.=S. Hassel, K.K.=K. Kristoffersen, L.L.=L. Lund, B.B.=B. Brunn, E.K.=E. Klüver, ?=analyst unknown.

Standard No.	Analyst	SiO_2 %	Al_2O_3 %	Total Fe as Fe_2O_3 %	CaO %	MgO %	K_2O %	Na_2O %	TiO_2 %	MnO %	
1	S.H.	63.45	16.18	5.51	3.54	1.00	4.60	3.85	0.97	0.10	
2	S.H.	61.35	16.53	6.54	3.74	0.99	4.28	3.90		0.12	
3	S.H.	66.6	15.12	4.10	2.30	0.62	6.23	3.40		0.06	
4	S.H.	64.27	18.50	0.57	0.40	0.52	11.70	4.30			
5	K.K.	61.68	18.01	4.98	5.50	1.42	0.98	4.96	1.18	0.13	
6	K.K.	70.05	11.64	6.02	3.05	2.69	0.31	2.95	0.78	0.09	
7	K.K.	47.76	13.07	15.24	10.10	6.25	0.37	2.37	3.28	0.23	
8	K.K.	42.47	6.15	14.22	0.94	26.40	0.10	0.00	0.65	0.12	
9	K.K.	54.34	18.26	11.11	0.16	5.61	3.83	0.35	1.00	0.12	
10	K.K.	48.32	24.36	9.50	2.73	4.16	4.75	1.86	0.95	0.10	
11	?	38.73	18.36	6.32	30.99					3.89	
12	?	35.44	1.26	29.63	32.62	0.0				0.48	
13	?	38.40	21.01	30.30	8.61	4.02				0.70	
14	?	39.29	21.70	34.3	1.99	5.26				1.51	
15	L.L.	42.40	18.22	7.01	19.72	0.83	0.03	0.14		4.42	B_2O_3 6.05
16	L.L.	51.55	15.44	9.66	8.17	8.71	1.43	2.82	0.68	0.06	
17	B.B.	69.25	14.65	4.67	2.32	1.44	3.15	3.38	0.42	0.04	
18	B.B.	74.53	13.64	2.18	0.27	0.05	4.70	3.98	0.01	0.01	
19	B.B.	68.45	15.45	4.46	2.50	1.27	2.60	3.90	0.36	0.05	
20	L.L.	49.65	13.67	10.64	13.53	6.63	0.83	3.38	1.39	0.37	
21	B.B.	63.56	15.04	7.18	0.70	0.60	6.73	3.03	1.08	0.13	
22	B.B.	7.1	2.2	1.9	46.8	3.4					
23	?	45.65	12.12	13.27	9.46	11.30	1.18	2.64	2.66	0.20	-
24	B.B.	45.0	16.34	14.76	9.15	5.63	2.15	3.20	3.3	0.68	
25	B.B.	53.46	$+P_2O_5$ 4.96	1.53	20.32	0.76	1.55	1.17	0.20	0.02	CO_2 15.25 H_2O 7.23
26	B.B.	63.54	23.89	1.02	0.11	0.25	2.14	0.41			
27	K.K.	28.40	7.30	0.64	35.48	18.00	0.24	0.26	9.04	0.18	
28	K.K.	11.10	1.89	28.52	43.63	4.82		0.10	0.31	8.22	WO_3 1.10
29	K.K.	21.10	6.03	3.40	64.42	3.39			0.31		
30	B.B.	41.21		7.87		49.41					
31	E.K.	43.34	32.81	0.45	0.56		4.63	16.23			
32		ca. 100									

* Nos. 1–25 are minerals and rocks [Nos. 1–10 are rocks described thoroughly by C. W. Carstens (Petrologische Studien im Trondhjemsgebiet, Det Kgl. Norske Videnskabers Selskabs Skr. 1928, No. 1; Norsk Geol. Tidsskr. 8, 235 (1925), Nos. 11–14 garnets, Nos. 23 and 24 diabases], No. 26 is clay, No. 27 steel reduction slag, No. 28 steel oxidation slag, No. 29 portland cement, No. 30 olivine rock, No. 31 nepheline, and No. 32 pure quartz.

For the determination of Si, the only internal standard that was found to be satisfactory was Be.

The internal standards were mixed into the samples in the form of carbonates. $SrCO_3$ was purified by converting it to nitrate, treating this with concentrated HNO_3 several times and converting to carbonate again.

The $BeCO_3$ used was dehydrated and practically Si-free.

DETERMINATIONS OF VARIOUS ELEMENTS

Table II shows the arcing conditions and the mixtures that must be made for the carrying out of a complete analysis.

The determination of Si was found to give best results when sample and internal standard were mixed with NaCl; as this mixture is quite different from those made for the other determina-

tions, Si must be determined in a separate operation.

K and Na. Another element that must be arced separately is K, as this element has no lines of sufficient intensity in the ultraviolet. The K-line used is 4044.1A and as this line lies in the cyanogen bands, which must be supressed, and since its sensitivity is not very high, there is nothing gained by arcing the sample as anode. Na is also best determined when arced as cathode; the spectrograph should therefore, if possible, be set so as to cover in addition part of the ultraviolet region down to 3300A, thus including the line pair Na 3302.3–Sr 3301.7. If this is not possible, Na should be determined by anode evaporation along with Al, Ca, Fe, etc.

The photographing of the spectrum in the visible region adds the opportunity of determining Ca simultaneously by the very good line pair Ca 4302.5–Sr 4361.7, which pair is especially suited for the determination of small amounts of Ca, and is also very good for large amounts of this element.

Al, Ca, Mg, Fe, Mn (and Na). The third spectrum is photographed in the ultraviolet region, for the simultaneous determination of Al, Ca, Mg, Fe, Mn, (and Na) and the composition of the mixture used may be the same as used for the K-determination. However, when some of these elements are present in large amounts, the determination is often carried out more conveniently with a diluted sample. The dilution is made by adding pure quartz to the sample as indicated in Table II, the $SrCO_3$ content being kept at 50 percent. The advantage of choosing quartz as a diluting compound is obvious: (1) As quartz is a usual constituent of rocks, the rock-quartz mixture still retains the nature of a rock powder. (2) Very pure quartz is fairly easily available and can be purified still more by treating the finely ground quartz powder with warm 10 percent HCl. (3) As SiO_2 in any case is determined in a separate mixture, the adding of quartz does not disturb the Si analysis in any way.

The burning of the sample as anode for this determination yields so many advantages that the cathode method cannot compete. As reported by Carlsson[2] and by many American spectrographers the anode method often gives a very good reproducibility, and the present work has also shown its superiority in this special field. The sample is also consumed in a much shorter time when arced from the anode, which is an advantage when speed is demanded. Upon arcing the actual rock $+SrCO_3$ or rock-quartz $+SrCO_3$ mixtures in the anode, many irregularities, which often occurred when evaporating the same mixtures from the cathode, were smoothed out and quite disappeared. The improvement was probably due to the high temperature of the anode, which results in the uniform melting of the substance and a rapid and very steady burning of the sample.

The larger size electrodes in Table II were used when greater intensity was required. In both

TABLE II. Arcing conditions.*

Spectrum	Size of lower electrode. (Upper electrode is of 4 mm diam.)	Sample placed in:	Part of arc column used. (Electrode gap = 10 mm.)	Amps.	Exposure time = arcing time	Proportions of mixture %
Visible and upper ultraviolet for K, Na (and Ca)	1½ × 4 mm crater. 3.0 mm outer diam.	Cathode	¾'s nearest cathode	6. no preheat	2 minutes	Sample +$SrCO_3$ 1:1
Ultraviolet for Si	1½ × 4 mm crater. 3.0 mm outer diam.	Cathode	¾'s nearest cathode	7. moderate preheat 10 sec.	2½ minutes	Sample + $BeCO_3$† +NaCl 1:1:3
Ultraviolet for Ca, Al, Fe, Mg, Mn (and Na)	1½ × 4 mm crater and 3.0 mm outer diam. or 2 × 6 mm crater, and 3.5 mm outer diam.	Anode	Middle ¾'s	7. moderate preheat 10 sec.	About 50 sec.	Sample + $SrCO_3$ 1:1 or sample +quartz + $SrCO_3$ 1:2:3

* To all mixtures 2 parts by weight of carbon powder is added.
† The use of the corresponding amount of BeO did not change the result.

[2] C. G. Carlsson, Jernkontorets Annaler **127**, 575 (1943).

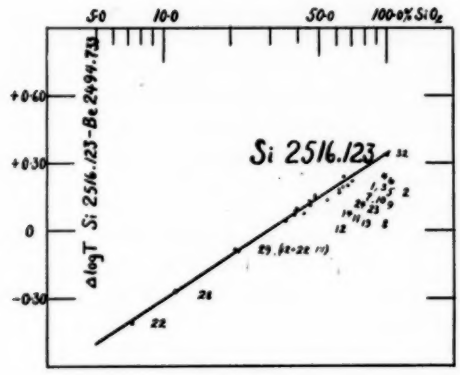


FIG. 1.

cases the substance was consumed in about 1 minute (50 seconds). This rapid evaporation is very convenient, as is also the appearance of the spectra produced, inasmuch as the latter are relatively weak and have almost no background. If spectra of still greater intensity are desired, they can be obtained by using electrodes of greater *outside* diameter, by which the evaporation time is prolonged.

DETAILS OF THE SPECTROGRAPHIC PROCEDURE

The electrodes used were made of Siemens carbons, purified according to the method described by Gatterer.[3] They were completely free from all impurities except boron.

The spectrographs were a quartz spectrograph constructed by R. Mannkopff in Göttingen, Germany, with a dispersion corresponding to 2.8–9.0 A/mm on the plates in the region used, and a two prism glass spectrograph also of Mannkopff construction, the dispersion being 4.1–6.0 A/mm. To save plate material during the war time, the ultraviolet spectra made by anode evaporation were cut off at about 2700A, so that the weak Al lines 2660–52A and 2575–68A could not be reached. One of these lines had to be used to complete the method, and this final part of the work was done in the laboratory of Spectrographic Services and Supplies Inc., the grating spectro-

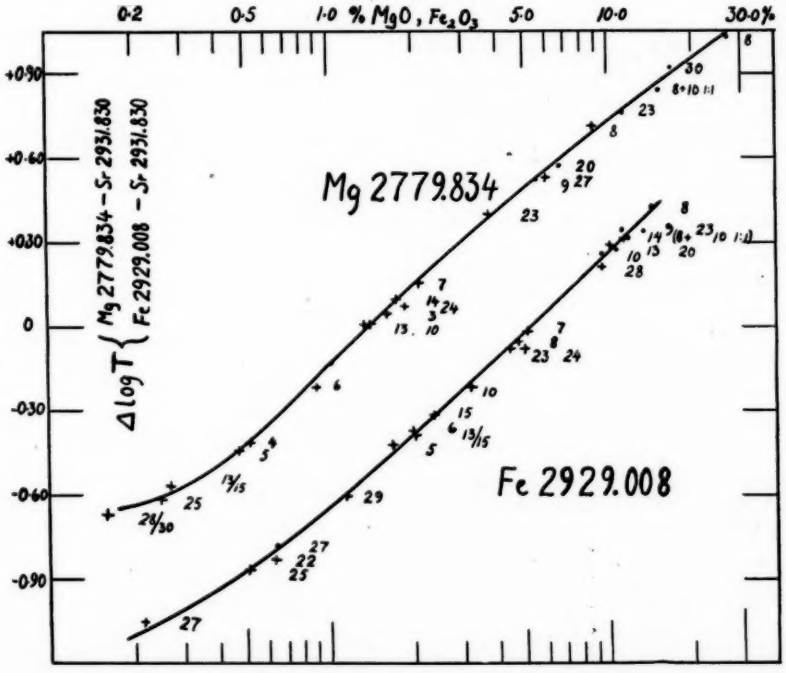


FIG. 2.

[3] A. Gatterer, Spectrochim. Acta 2, 49 (1941).

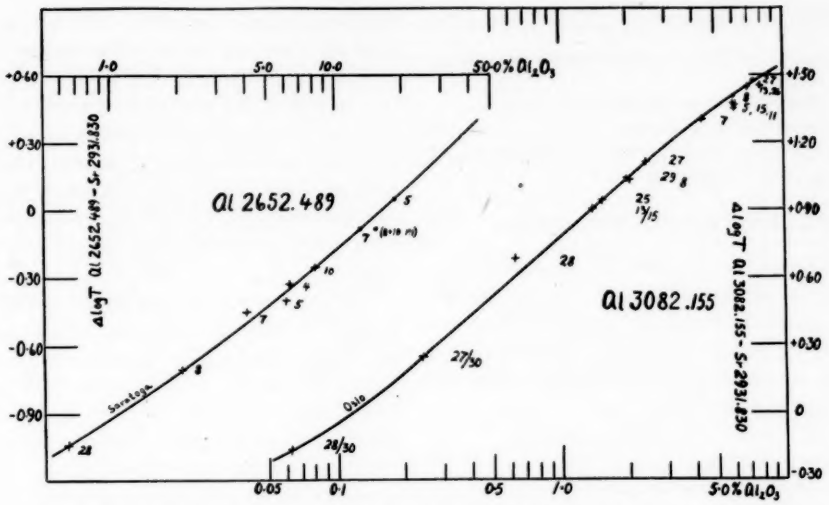


FIG. 3. Right: Calibration curve for the line pair Al3082–Sr2931 valid from about 0.05 percent to about 8 percent (Oslo spectrograph). Left: Calibration curve for the line pair Al2652–Sr2931 for large Al-amounts. (Saratoga Springs grating spectrograph.)

graph (from Harry S. Read Company) of this laboratory being simultaneously calibrated for this rock analysis method. The dispersion of this instrument was 3.3 A/mm in the second order.

The spectra were photographed on "Perutz Graphische Platte *B*" or "Perutz Reproduktionsplatte" (ultraviolet) and "Agfa Normal Film" (visible region), and developed in a Metol-Hydroquinone Developer. (The spectra made in Saratoga Springs were photographed on Eastman Spectrum Analysis Nr. 1 plates and also on Eastman 103–O plates).

For all spectra a step sector was employed, the analysis and comparison lines photometered and density curves plotted in the usual way.[4] Calibration curves were constructed as shown in the Figs. 1–8. The wave-lengths accompanying the curves are taken from Harrison's table.[5]

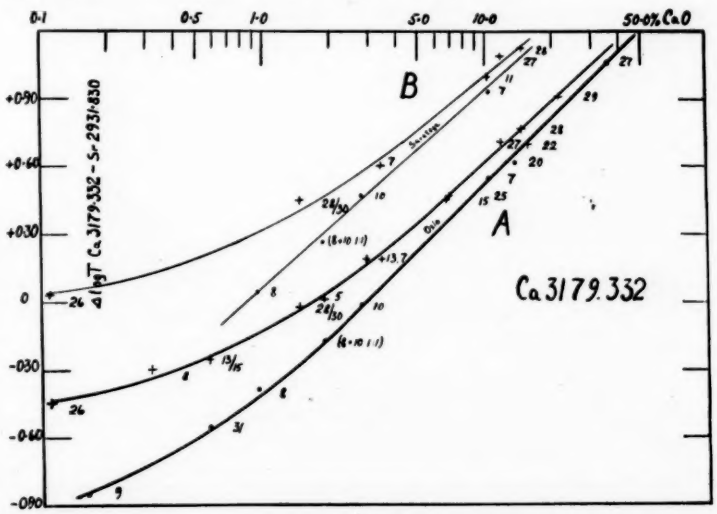


FIG. 4. Right: (curves *A*): Calibration curves for Oslo spectrograph. Left: (curves *B*): Calibration curves for grating spectrograph, Saratoga Springs. (+): quartz-diluted sample +$SrCO_3$. (·): undiluted sample +$SrCO_3$.

[4] See, e.g., L. W. Strock, A.I.M.E. Tech. Pub. No. 1866; or A. Kvalheim, Norsk Geol. Tidsskr. 21, 245 (1941).
[5] George R. Harrison, *M.I.T. Wave-Length Tables* (John Wiley and Sons, Inc., New York, 1939).

Explanation to the figures:

$\Delta \log T$ = density curve separation, measured on log relative exposure time scale.
A dot (·) means: Sample not diluted with quartz.
A cross (+) means: Sample diluted with 2 parts of quartz.
A cross and a fraction, e.g., +27/30 means: Sample No. 27 has been diluted 30 times with quartz.

SOME COMMENTS TO THE CURVES

Some interesting features of the calibration curves may be mentioned here. In spectrographic work it is often necessary to construct separate calibration curves for different electrode sizes when using the same line pair. When using the anode as carrier of substance in the rock analysis described, the calibration curves for Ca, Fe, Al, Mg, Mn, are the same for $1\frac{1}{2}\times 4$ mm carbon craters and for 2×6 mm craters, except for the upper part where absorption causes a separation.

These features of the calibration curves show that the substance quantity used may vary considerably without influencing the analysis at all, a fact which is very valuable to the spectrographer, since it eliminates a source of error that must otherwise often be taken into account.

For Fe, Al, Mn and Mg the curves are identical when using the 1:1 or the 1:2:3 mixture ratios; even 15 and 30 times dilutions are used, and the points representing the diluted samples still do not depart from the curve but mingle with the dots representing undiluted samples.

As to the Ca curves, it will be seen that the Ca line 3179A is influenced by the SiO_2 content of the sample, as separate curves must be applied for the 1:1 and for the 1:2:3 mixtures; however, such an effect cannot be detected on the other Ca calibration curve (Ca 4302), as the points from both mixture ratios in this case all fall on a single curve.

The exposure time for the determination of K and Na (Ca) was 2 minutes, which means that the exposure was cut off before the sample was totally consumed. (The sample will usually be completely evaporated in $2\frac{1}{4}$ to $2\frac{1}{2}$ minutes.) This limiting of exposure gives better results than are obtained by burning the sample to complete evaporation. It is also very valuable for small K-amounts as the background is kept fairly clear without weakening the K line (K 4044.1) which disappears at about 0.3 percent K_2O.

The simultaneous determination of Ca was found to be uninfluenced by this cut-off procedure. K was previously determined by the line pair K 4044.1A–Sr 4161.8A,[6] but the pair K 4044.1A–Sr 4032.4A was later found to be much more stable.

As mentioned before, the spectra of the weak Al-line 2652.5A was taken with a different

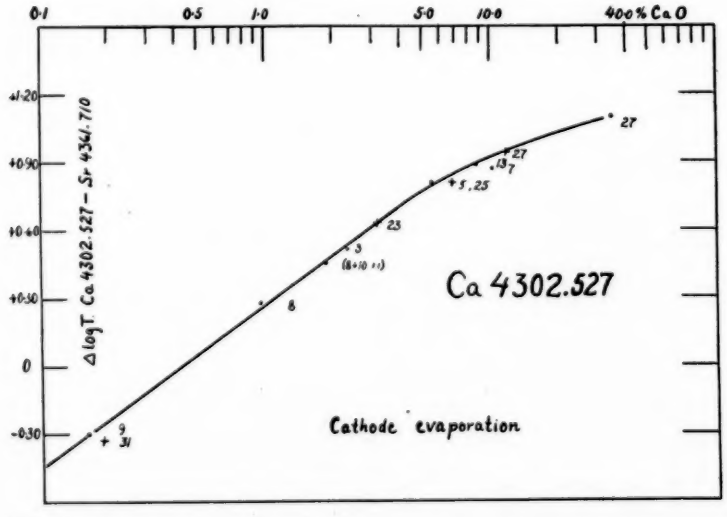


FIG. 5.

[6] A. Kvalheim, Fra Fysikkens Verden, Hefte **3**, 161 (Oslo, 1944).

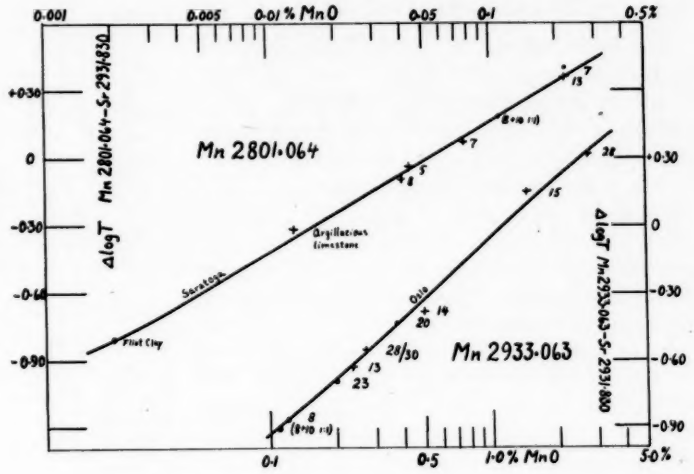


FIG. 6. Right: Calibration curve for Mn (0.1—about 5 percent MnO) (Oslo spectrograph). Left: Calibration curve for small amounts of Mn (grating spectrograph, Saratoga Springs). The two lowest points are flint clay and argillaceous limestone, Bureau of Standards No. 97 and 1A, respectively.

spectrograph (grating) than that used in obtaining the other spectra, and the calibration curve is thus not directly comparable with the other curves.

The Na calibration curve for *cathode* evaporation is also constructed only for this spectrograph; for Mn 2801, the author has chosen to present the curve obtained in Saratoga Springs, because the use of some new standards, e.g., Flint Clay, Bureau of Standards No. 97, prolonged the calibration curve down to 0.002 percent MnO.

An example of the difference between the two cases is shown for Ca in Fig. 4. The arcing conditions are *supposed* to be approximately the same, but different makes of plates were used.

THE INFLUENCE OF K ON THE DETERMINATION OF OTHER ELEMENTS

Only one element is found to be able to bring some sort of irregularity into the analysis of common rocks, *viz.* K. When the K content of the rock or mineral is higher than about 8 to 10 percent K_2O, the Al determination *may* give results which are more than 10 relative percent *too low*. It is therefore recommended, when a sample is found to contain much K, to dilute it, e.g., five times with quartz and check the result of the Al analysis. The other determinations should be checked simultaneously, since the enhancement of the Sr-lines, because of K, is not always accompanied by a corresponding change in the intensities of other lines.

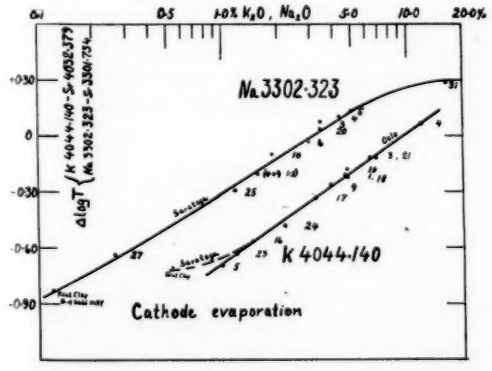


FIG. 7.

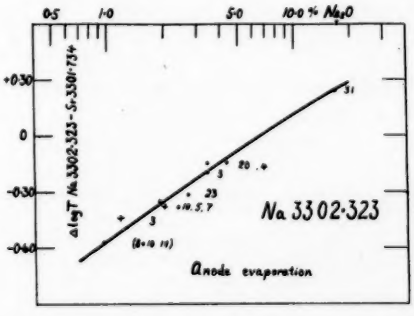


FIG. 8.

THE DETERMINATION OF Ti

The independence of the determinations described, with regard to changes in chemical combinations of the elements in question, and to variations in base substance, does not apply correspondingly when trying to analyse Ti.

It is often assumed that the addition of large amounts of one or more elements (for instance, alkalies) to the sample will eliminate the influence of differences of the kind mentioned above, but the benefit of this "choking" procedure is in many cases overestimated.

When constructing the calibration curve for Ti (line pairs Ti 3088.025–Sr 2931.830, and Ti 3241.980–Sr 2931.830) a line that was straight between 0.1 and 10 percent was the result, and all the points except two were very close to this curve. The two exceptions (Nos. 8 and 10) gave figures departing about −30 percent and +30 percent, respectively, from the true values when no quartz was added, and when the same samples were diluted with 2 parts of quartz the deviations were the same, about −30 percent for sample No. 8 and +30 percent for No. 10.

This shows that "choking" will not help in this case. The deviations are most probably due to the different nature of the Ti-carrying minerals present in those rocks. All the points on or close to the curve are represented by Ti in the form of rutile (TiO_2) or titanite ($CaTiSiO_5$); the two exceptions carry Ti in a biotite and in the form of ilmenite ($FeTiO_3$), respectively.

The case seems to be similar to that of Zr as reported by Strock.[7]

ERRORS AND REPRODUCIBILITY

As actual rocks are used as standards and not artificial mixtures of the oxides, the average errors usually involved in this analysis can be found directly from the Figs. 1–8. Each point (dot or cross) in these figures represents from 2 to 4 separate exposures, and it will be seen that the relative average error very seldom exceeds 10 percent (exception: Na when determined by anode evaporation). As a matter of fact the error is in many cases less than 5 percent.

To show the reproducibility of the method, a typical repeated shale analysis from Spectrographic Services and Supplies, Inc., is indicated in Table III where the values obtained from repeated determinations of Al, Ca, Fe and Mg in Flint Clay (Bureau of Standards No. 97) are also listed.

TABLE III.

	% Al_2O_3	% CaO	% Fe_2O_3	% MgO	% MnO	% K_2O	% Na_2O
Shale	16.0	0.70	7.8	1.30	0.11	1.60	0.78
	15.3	0.72	8.2	1.19	0.12	1.45	0.89
	15.6	0.72	8.0	1.22	0.11	1.30	0.68
Flint Clay	38.5	0.12	0.94	0.28	*		
(Bureau of	38.5	0.12	0.91	0.24			
Standards	42.0	0.14	1.05	0.23			
No. 97)	40.0	0.13	0.97	0.25			

* The Mn, K and Na-figures of this clay are used as lowest points on the calibration curves for these elements.

ACKNOWLEDGMENT

The author is indebted to the Norwegian Ministry of Commerce for the opportunity of carrying out this work, and to Dr. Lester W. Strock for being allowed to do some final work on the method in the laboratory of Spectrographic Services and Supplies, Inc.

Professor M. Raeder; B. Bruun, chemical engineer; K. Kristoffersen, curator; and L. Lund, M.A., have all supplied chemically analyzed samples for this work, for which courtesy the author wishes to express his gratitude. Sincere thanks are also due Mrs. Else Moursund for valuable assistance and participation in this work.

[7] L. W. Strock, A.S.T.M. Bulletin, **109**, 37 (March, 1941).